

# THIS WEEK

## EDITORIALS

**WORLD VIEW** Why nuclear power must survive as an energy source **p.411**

**CELL BIOLOGY** X-rays capture first drug binding to receptor protein **p.412**



**TITAN** Rain on moon's plains is mainly methane **p.413**

## The long road back

*For now, Japan's scientists have higher priorities than rebuilding their research infrastructure, but when they do get to it, they will need help from the international scientific community.*

The Japanese prime minister, Naoto Kan, has called this month's earthquake, tsunami and devastation at a nuclear-reactor plant the worst disaster to hit the country since 1945. The death toll has soared past 20,000, and the full extent of the damage won't be known for some time (see News page 420).

The situation also been a catastrophe for science in Japan. Laboratories have been destroyed, and Tohoku University in Sendai, one of Japan's premier research institutions, will be closed until at least the end of next month. Many buildings can't be entered, and broken equipment and destroyed samples fill those that can. The impact stretches down the east coast to Tsukuba Science City, where 40% of the country's researchers are based. Even in the greater Tokyo area, where most facilities escaped physical damage, much research has come to a halt because of blackouts and an exodus of expatriate researchers, who have left because of worries about radiation.

Alongside the humanitarian aid that has poured in from dozens of countries, scientists around the world are offering to help their Japanese colleagues. Some have approached acquaintances, whereas others are taking more formal approaches.

The US National Institutes of Health is planning to provide temporary scientific homes for those who have lost research facilities in Japan. The Nippon Science Support Network, supported by Nature Network, is helping to coordinate scientific relief efforts from Germany; as of Tuesday, the site had 18 offers of scientific positions and other services such as computer-server space, many of them fully funded, ranging from mathematics to molecular pharmacology and plasma and astrophysics. An international grass-roots initiative is collecting small-scale support in the form of accommodation, funding, lab space and server space. The German national academy of sciences, the Leopoldina in Halle, the German Academy of Science and Engineering in Berlin, and the Berlin-Brandenburg Academy of Sciences and Humanities have offered €5 million (US\$7.1 million) to support Japanese science. A group at the Chinese National Center for Nanoscience and Technology in Beijing has offered to host scientists. And institutions within Japan are discussing loans or donations of instruments.

Those making these generous offers shouldn't be surprised if they are not taken up just yet. Many affected scientists don't have consistent access to the Internet, and most are more concerned with the necessities of daily life. Before going abroad, scientists are trying to work out what they can salvage of their laboratories at home. And although the best thing for their research might be to move to a facility where basics such as water and electricity are not a problem, many — especially senior researchers — have work or family obligations in Japan.

This will change as researchers size up what they need to do to reconstruct their research. The best thing the scientific community can do is to keep these offers open. Young Japanese researchers, especially, should be ready to take advantage, and this could be a timely boost for the

nation's science: in the past ten years, the number of young researchers travelling abroad has fallen dramatically, producing an insular research community that could benefit from more outside contact.

The destruction also brings research opportunities. The devastation shows the power of tsunamis and earthquakes and the seriousness of energy shortages, and associated medical problems could bring home

***"Before going abroad, scientists are trying to work out what they can salvage at home."***

the importance of science to a generation of uninterested school children, whose curriculum currently contains less science than at any time in recent history. Ryoichi Matsuda, a biologist at the University of Tokyo, suggests that the tragedy could be used to re-emphasize "science education for survival".

Rebuilding also has its benefits. Scrapping of old nuclear reactors will open discussions of other options, such as the introduction of geothermal energy. Tohoku University administrators are talking about improving the institution's infrastructure, which could see wider refurbishment of outdated facilities.

In a time of such death and devastation, the scientific infrastructure will, of course, not be the country's first priority. Scientists throughout Japan are preparing for cuts to help the northeast get back on its feet. But the nation cannot survive without science and technology.

The Japanese government will, no doubt, step up to the challenge of rebuilding its science, but there will be a long struggle to build a solid foundation, and many research lives could fall through the cracks. Those who are creating windows of opportunity for Japan's needy scientists should keep them open. And others might want to think about opening more. ■

## Contact your MP!

*British readers should help to change libel laws that suppress global scientific discussion.*

Britain's Houses of Parliament were earlier this month the scene of a gathering by an array of campaigning and media organizations, who came together to press for reform of the libel laws of England and Wales. Amnesty International UK, Global Witness, Facebook, Mumsnet, the *British Medical Journal* and *Nature* were among those represented, urging a group of Members of Parliament (MPs) to support the introduction of new legislation.

The few dozen MPs at the meeting needed little persuading — but there are more than 600 others, and more proposed legislative issues

than can be handled within the current Parliament, before the next general election. So the fight for the attention of MPs will be crucial in the coming months. It is a fair bet that few of them, if any, have ever heard from constituents on this issue. But they do pay attention to their mailboxes, and there is now both an opportunity and a need for anyone concerned about these issues to help to ensure that essential legislative reform is seen through.

The good news is that all three of Britain's main political parties support such reform. Over the years, it has become increasingly clear that the burden of libel litigation falls too heavily on those who write about misconduct and bad practice. And with the global nature of the Internet, anyone outside the United Kingdom can sue anyone else outside the country using this law, provided that the libel was accessible to readers in the United Kingdom. Journalism on scientific issues has been acutely affected, and the need for reform and examples of problems within science were highlighted in these pages last year (see *Nature* 464, 1104; 2010).

At *Nature*, we have too often been hindered in our core mission because of legal risks. On one occasion we were unable to link to a university's website to point our readers to the outcome of a misconduct investigation, associated with the retraction of a research paper, because of a threat from the person found guilty by the university. There has been journalism about misconduct — central not only to the interests of *Nature's* readers but also to public trust in science — that we have decided not to commission, because we decided that the risks of costly libel action outweighed the undoubted significance of the stories. We will always pursue the most significant cases of transgression — on one occasion at very considerable legal expense. But there is a layer of less egregious yet still significant misconduct that we are not covering because of the risks of such costs.

Britain's coalition government has now introduced draft legislative

reform that would allow us to perform our core mission with fewer restrictions. Part of the problem with the existing libel laws is that they place a heavy burden of proof on the defendants and have little scope for a public-interest defence — areas that will be retuned under the proposed reform.

Ministers responsible for the proposed changes specifically mention the freedom of scientific debate as one focus of their concern.

***"The attention of Members of Parliament needs to be sharpened now."***

The changes and a consultation paper can be found at [go.nature.com/o3vw5r](http://go.nature.com/o3vw5r).

We at *Nature* will respond to that consultation. And we urge readers to respond to the detailed questions if they are seriously interested in strengthening the ability of scientists, journalists and others to report and comment publicly on misconduct or to speak responsibly and freely about problematic products or actions of large institutions and companies. The Libel Reform Campaign — a coalition of interested organizations — has published an initial response to the draft bill making clear their view of what further changes should be sought (see [go.nature.com/vdjvna](http://go.nature.com/vdjvna)).

The consultation process has a deadline of 10 June 2011. The bill will then be amended and formally introduced to Parliament for implementation in 2012. That will be another moment at which support for the bill will be crucial. But the attention of MPs needs to be sharpened now. To that end, readers in the United Kingdom should immediately contact their MPs to draw attention to the issue and to urge their support. The organization Sense About Science has published a template letter and MPs' contact details — see [go.nature.com/xrdcfx](http://go.nature.com/xrdcfx) and ask MPs to sign Early Day Motion 1636, tabled by Cambridge MP and scientist Julian Huppert and supported by cross-party colleagues. ■

## A unifying cause

*Conference of science journalists can strengthens ties between the Arab world and the West.*

With the recent awakening in the Arab world of movements for democracy and free speech, it is timely that the World Conference of Science Journalists (WCSJ), on 27–29 June, will for the first time be held in an Arab country. Even organizing the conference in Qatar has, in its own small way, promoted collaboration between the Western and Arab journalists involved. It can only be hoped that the mingling of science reporters at the event will have a similar, and lasting, effect. Western journalists attending the conference should take the opportunity to see the Middle East, meet its scientists and learn more about how science might contribute to sustainable development of the region, and the substantial challenges it faces, in particular at this crucial and historic moment in the region's history. Support for science in the Arab world has long been at levels far below those in other countries, although there have been some recent improvements (see *Nature* 470, 147–149; 2011). A twinning between the young Arab Science Journalists Association and the well-established US National Association of Science Writers in 2007 made the joint bid to bring the conference to the Arab world possible. That twinning also built powerful ties between science journalists in the Arab world and in the United States. Arab journalists were invited to American science and science journalism conferences, and American journalists attended the first Arab science journalists conference in 2008. There was much to learn for both sides as they shared challenges, advice and opportunities. It created mutual understanding between two regions that are often perceived as being at odds with one another.

It is a great pity, although understandable given the recent unrest and uncertainty in Egypt, that the organizers decided to relocate the conference from its original planned venue in Cairo to Doha in Qatar. It would have been symbolic to hold a major conference of journalists in a nation that has just overthrown the shackles of a dictatorship that repressed free speech and the critical thought and questioning that science and science journalism thrive on. But at least the venue has been kept in the Arab world, and has not been moved to the United States, which was discussed as an alternative venue at one point.

Holding the conference in Qatar will hopefully also provide a boost to science journalism in the region, which has suffered, as has all journalism and civil society, under authoritarian regimes. There are no dedicated science journalism courses in any of the universities in Arab states and, although there have been improvements, much of the science journalism there remains poor. The conference is a chance for Arab science journalists to rub shoulders with colleagues from all over the world and exchange their experiences. The connections made will be invaluable as science becomes more global. Many local and regional organizations are now thinking about projects they can put together to train and support science journalists. This will create a momentum to support the profession long after the conference has come and gone.

Past conferences have catered too much to Western issues, but this year's WCSJ, with a rich programme and speakers from more than 40 countries, promises to begin providing greater balance. Speakers from the Arab World, Africa, Latin America and Asia will give delegates greater insights into the science needs and challenges of the developing world. There is much reconstruction of civil society to do in the fledgling democracies of Tunisia and Egypt, and

science journalism can play its own small part in prompting debate on crucial science-based issues in every sector, as well as bringing greater scrutiny to the glaring needs in research and higher education. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunjv](http://go.nature.com/xhunjv)





## Do not phase out nuclear power — yet

*Fission power must remain a crucial part of the energy mix until renewable energy technologies can be scaled up, argues Charles D. Ferguson.*

The ongoing Japanese nuclear crisis underscores yet again the risks inherent in this essential energy source. But it should not divert nations from using or pursuing nuclear power to generate electricity, given the threat from climate change, the health hazards of fossil fuels, and the undeveloped state of renewable energy. Instead, the events at the Fukushima Daiichi Nuclear Power Plant should turn more attention to ensuring that nuclear power plants meet the highest standards of safety and protection against natural disasters.

More than 30 nations have commercial nuclear power plants. A further two dozen are interested in having them, including several in earthquake risk areas such as Indonesia, Malaysia and Turkey.

Some nations are pro-nuclear for energy security; some for prestige. Others, including Iran, have invested in nuclear power because they may want the capability to make nuclear weapons.

These nations are seeking to acquire uranium enrichment or reprocessing technologies: useful either for producing fuel for peaceful nuclear reactors or fissile material for nuclear bombs.

Although some national leaders profess to be interested in nuclear energy because operating plants do not emit greenhouse gases, this is usually a secondary motivation. If it were their primary concern, nations would invest far more than they have in measures such as energy efficiency and solar and wind technologies.

The Japanese crisis has affected three important criteria: public opinion, safety and economic costs. Governments and utilities have had to grapple with these for decades. Now they must renew their efforts to finance expensive nuclear projects and ensure that existing and future nuclear plants maintain the highest standards — and must be seen to do so by the public.

Building nuclear power plants has always been expensive. For a large reactor with a power rating of 1,000 megawatts or greater, the capital cost ranges from US\$4 billion to \$9 billion depending on reactor design, financing charges, the regulatory process and construction time. The recent nuclear crisis is likely to change all of these, pushing up costs.

Contemporary plant designs — 'generation III' — have better safety features than the 1970s-era generation II designs for the Fukushima reactors, making them more expensive. Some, such as the AP1000 designed by the Westinghouse Electric Company, headquartered in Cranberry Township, Pennsylvania, have passive safety features that do not require technicians to activate emergency systems or electrical power to ensure safety after a mishap. Others, such as Paris-based Areva's EPR, have advanced active safety systems designed to prevent the release of radioactive material to the environment. Further designs, such as the pebble-bed modular reactor, may prevent nuclear fuel from ever experiencing

a meltdown. Concerns were raised about the Fukushima designs as early as 1972, the year after reactor unit 1 began operations. But the nuclear industry opposed shutting down such reactors because 32 were in operation worldwide — about 7% of the world's total. Almost one-quarter of the reactors in the United States are of this type. The remaining plants of this design should undergo a thorough safety review and, as a result, some may need to close. Since the crisis began, several governments, including China, Germany and Switzerland, have called for increased scrutiny of their plants and a moratorium on plant construction until plant safety is assured. Germany has also shut down its seven oldest reactors.

But phasing out nuclear power worldwide would be an overreaction. It provides about 15% of global electricity and even larger percentages in certain countries, such as France (almost 80%) and the United States (about 20%). Eliminating nuclear power would lead to much greater use of fossil fuels, and raise greenhouse-gas emissions. It will probably take at least a few decades to massively scale up use of renewable sources. Meanwhile, nuclear plants can bridge the energy gap.

So governments need to take practical actions to improve nuclear safety. All new nuclear plants should have enhanced safety systems, and plant designs that eliminate or substantially reduce the risk of a meltdown of fuel should be developed. Existing plants deemed to fail improved safety standards should be retrofitted or, when necessary, phased out. Further, governments must force their nuclear providers to remove spent fuel — typically after five years of cooling — from storage pools and place it in dry cask storage. As the world witnessed, spent fuel in the overcrowded above-ground cooling pools at Fukushima Daiichi became exposed to the air. If spent fuel catches fire, radioactive materials can be widely dispersed.

Because of decreased public confidence following the Japanese accident, governments and industry must have an honest conversation about the role of nuclear energy in meeting consumers' electricity demands, the typically high safety record of almost all plants and the risks of this technology. These discussions must implement one of the primary lessons of the Japanese accident: that officials should dramatically increase transparency of nuclear operations. Simultaneously, nations need to invest far more in renewable energy sources, which offer the path to a truly sustainable global energy system. ■

**GOVERNMENTS AND  
INDUSTRY MUST HAVE  
AN HONEST  
CONVERSATION  
ABOUT THE ROLE  
OF NUCLEAR  
ENERGY.**

**➔ NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/jjm47y](http://go.nature.com/jjm47y)

**Charles D. Ferguson** trained as a nuclear engineer and a physicist, and is president of the Federation of American Scientists in Washington DC and author of the forthcoming book *Nuclear Energy: What Everyone Needs to Know* (Oxford University Press, May 2011). e-mail: [cferguson@fas.org](mailto:cferguson@fas.org)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## CANCER

### Liver cancer lifeline found

A survey of liver tumours has highlighted a gene that many such tumours depend on for survival.

Scott Lowe and Scott Powers at Cold Spring Harbor Laboratory in New York and their colleagues searched the genomes of 89 liver tumours and 12 liver-cancer cell lines. They identified 124 genes that were sometimes expressed in excess; overexpression of 18 of these caused liver cells transplanted into mice to become cancerous.

In particular, cells overexpressing the gene *FGF19* became dependent on this expression. An antibody that blocks the FGF19 protein inhibited the growth of these cells, suggesting that patients in whom this gene is overexpressed could benefit from therapies that block the protein.

**Cancer Cell** doi:10.1016/j.ccr.2011.01.040 (2011)

## CELL BIOLOGY

### Plot twist for proteins

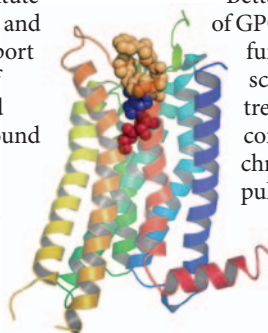
Cells interpret external chemical signals through membrane-spanning receptors that bind the chemicals and change shape to alter the cells' activities. Raymond Stevens at the Scripps Research Institute in La Jolla, California, and his colleagues now report the structure of one of the G-protein coupled receptors (GPCRs) bound to an activating drug.

The authors used X-ray crystallography to reveal the shape of the A<sub>2A</sub> adenosine GPCR (pictured)

bound to an agonist called UK-432097. This is the first time an agonist has been shown to bind to and stabilize the receptor without the aid of a G protein.

Better understanding of GPCR structure and function could help scientists to develop treatments for conditions such as chronic obstructive pulmonary disease.

**Science** doi:  
10.1126/  
science.1202793  
(2011)



## ECOLOGY

### Better fragmented than lost

Separating the effects of habitat loss and habitat fragmentation is difficult. To solve this problem, Mary Bonin and her colleagues at James Cook University in Townsville, Australia, arranged a series of experimental reefs off Papua New Guinea.

Few damselfish survived on reefs of *Acropora subglabra* (pictured) from which 75% of coral had been removed, whereas those on reefs that had been broken up but maintained in area

actually did better than those on untouched control reefs. Species richness and abundance were also higher on fragmented reefs than on those that had lost habitat. Although the positive effect of fragmentation declined over a 16-week period, the impact of habitat loss worsened in this time, suggesting that reported declines in fish populations after habitat disruption are down to the latter and not the former.

**Ecology** doi:10.1890/10-0627.1 (2011)

## PHYSICS

### An 'electric' force for neutral atoms

Neutral atoms can be made to behave like charged particles by 'synthetic' electric and magnetic fields. These are created through the production of a synthetic gauge field in a state of matter known as a Bose–Einstein condensate (BEC), in which the atoms are all identical and behave collectively as if they were one 'superatom'.

Ian Spielman and his team

at the National Institute of Standards and Technology in Gaithersburg, Maryland, have previously generated synthetic magnetic fields by spatial alteration of a time-independent electromagnetic vector potential — an entity that can be used to specify both the electric and the magnetic fields. Now they show that a synthetic electric field can be generated in a rubidium BEC in a parallel manner — by changing a spatially uniform vector potential over time.

The neutral atoms in the condensate were accelerated

like charged particles in the synthetic electric field that was induced by changing the electromagnetic vector potential.

*Nature Phys.* doi:10.1038/nphys1954 (2011)

## PLANETARY SCIENCE

## Methane rain falls on Titan

Images from NASA's Cassini probe revealed vast lakes of liquid hydrocarbons around the poles of Titan, Saturn's largest moon, in 2006. Elizabeth Turtle of Johns Hopkins University's Applied Physics Laboratory in Laurel, Maryland, and her colleagues now add the discovery that methane probably rains on Titan at low equatorial latitudes.

Images of the arid equatorial region taken by Cassini last October show dunes that appear darker after clouds had passed overhead. This suggests that the clouds rained liquid methane, which wet the surface, says the team. Such rain is thought to be seasonal and may play a part in dune formation by cementing fine atmospheric aerosol particles.

*Science* 331, 1414–1417 (2011)

## CHEMISTRY

## Copper makes for selectivity

Carbon-based compounds attached to a lithium atom are widely used in chemical reactions to add new carbon–carbon bonds to molecules. However, these organolithium reagents are not normally selective enough to create only one of two possible mirror-image forms, or enantiomers, of the same molecule.

Ben Feringa, Syuzanna Harutyunyan and their colleagues at the University of Groningen in the Netherlands now show off a way to use

these organolithium reagents in a highly selective manner. Using these reagents, the team demonstrated selectivity of 90% or higher for one enantiomer in more than 20 different reactions.

*Nature Chem.* doi:10.1038/nchem.1009 (2011)

## DEVELOPMENT

## How the chicken's neck got naked

How chickens of a particular breed ensure that their necks remain feather free — an adaptation to hot weather — has been explained.

Denis Headon at the University of Edinburgh, UK, and his co-workers found that naked neck chickens (pictured) have a mutation that boosts levels of the protein BMP12. Adding this protein to neck skin tissue cultures from normal chickens blocked feather formation, although the treatment had little effect on skin cultures from other parts of the body. The skin on chickens' necks naturally contains high levels of a signalling molecule called retinoic acid. This, the researchers show, primes neck skin cells to respond to the extra BMP12 made by naked neck chickens, preventing feather growth.

*PLoS Biol.* 9, e1001028 (2011)

## CLIMATE MODELLING

## Melting of the third pole

Aerosols such as black carbon and dust particles seem to have a greater effect on the Tibetan Plateau's snow than does anthropogenic climate change.

Yun Qian of the Pacific



## COMMUNITY CHOICE

The most viewed papers in science

## COMMUNITY CHOICE

## Skeleton boosts stud quotient

**HIGHLY READ**  
on [www.cell.com](http://www.cell.com)  
17 Feb–16 Mar

That gonadal hormones influence bone remodelling has been well documented, but it is not the end of the story. Gerard Karsenty at Columbia University in New York and his colleagues have found that this influence runs in both directions — although only in males.

They show that the hormone osteocalcin, made by bone cells called osteoblasts, induces testosterone production by testicular Leydig cells, the body's key testosterone factories. The researchers demonstrate this in both mouse-cell cultures and live mice, and also identify an osteocalcin receptor expressed in Leydig cells but not in follicular cells of the ovary. Male mice engineered to lack this receptor are subfertile, as are male mice engineered to lack osteocalcin.

*Cell* 144, 796–809 (2011)

## MICROBIOLOGY

## Sugary coat of armour

The soil-dwelling bacterium *Bacillus cereus* is a close relative of the microbe responsible for anthrax, and can cause a similar illness if it is inhaled by people with damaged lungs.

Olaf Schneewind and his co-workers at the University of Chicago in Illinois find that a pathogenic strain of *B. cereus* harbours two sets of genes that encode protective sugar coats. These coats prevent the bacteria from being engulfed and destroyed by certain immune cells. One gene set produces a protective capsule made of the sugar hyaluronic acid, whereas the other capsule is made of an unidentified sugar.

Knocking out one set of the capsule-making genes reduced the microbe's virulence in mice; if both were knocked out, the bacteria no longer caused disease.

*Mol. Microbiol.* doi:10.1111/j.1365-2958.2011.07582.x (2011)

**NATURE.COM**

For the latest research published by Nature visit:

[www.nature.com/latestresearch](http://www.nature.com/latestresearch)



# SEVEN DAYS

The news in brief

## POLICY

### UK libel reform

Long-awaited reforms to English and Welsh libel laws were proposed in draft legislation published by the UK government on 15 March. Changes to the existing laws could establish greater legal protection for scientists and journalists wanting to debate scientific or medical issues. The proposals suggest explicit protection for those reporting on academic conferences. A consultation runs until 10 June. See page 409 for more.

### US budget limbo

Researchers in the United States face another anxious wait before finding out the extent of cuts to science agencies' funding in the federal budget for the 2011 fiscal year. On 18 March, President Barack Obama signed a short-term continuing resolution to fund the government until 8 April, avoiding a shutdown. Until then, the Republican-majority House and the Democrat-majority Senate will continue to thrash out their differences on 2011 budget cuts (for details, see *Nature* 471, 144–145; 2011).

## SOUNDBITE

“We’re on the verge of an Arctic ozone hole.”

**Markus Rex of the Alfred Wegener Institute for Polar and Marine Research in Potsdam, Germany. See [go.nature.com/pkzgs1](http://go.nature.com/pkzgs1) for more on the Arctic’s unprecedented ozone depletion this spring.**



REUTERS/R. ORLOWSKI

## Fukushima sparks anti-nuclear protests

Japan’s nuclear crisis has shaken public confidence in the safety of atomic energy, but most governments are reluctant to pull the plug on nuclear plans. Instead they are promising to rethink strategies and undertake safety checks. European Union leaders decided that existing reactors in the region should undergo stress tests, which are voluntary. Facing public protests (pictured), Germany temporarily switched off its seven oldest reactors and put a three-month

moratorium on plans to extend reactor life-spans. Although Britain, Russia, France, the United States and India all announced safety checks of existing plants last week, none is delaying expansion plans. China, however, announced plans to temporarily suspend work on 26 reactors under construction, and to reconsider long-term expansion plans involving at least 50 more reactors. See [go.nature.com/sm6blc](http://go.nature.com/sm6blc) and page 411 for more.

### Elite spared cuts

Research-intensive universities in England will be cushioned from cash cuts to the nation’s higher-education grants. Compared with 2010–11, institutions next year will lose on average 3.7% of their research and teaching grants, which respectively total £1.6 billion (US\$2.6 billion) and £4.3 billion. But provisional allocations released on 17 March show that small, teaching-focused institutions will lose up to 10% in grants, whereas research-intensive universities such as Cambridge are to see budget cuts of less than 1% — and

some, such as Oxford, will see increases. See [go.nature.com/2xgqfj](http://go.nature.com/2xgqfj) for more.

## BUSINESS

### Celera sold

Pioneering genetic sequencing company Celera has been sold to the medical giant Quest Diagnostics for US\$671 million, it announced on 18 March. In 1998, Celera boasted that it would sequence the human genome within three years — and then succeeded. Since then the company, founded by geneticist Craig Venter and based in Alameda, California,

has focused on developing genetic tests for diseases. The deal will give Quest Diagnostics, of Madison, New Jersey, access to Celera’s pipeline of disease biomarkers.

### Solar subsidy cuts

The United Kingdom has signalled its intent to join Spain, Germany, France and Italy in reducing state incentives for solar-power production. In a consultation document published on 18 March, the UK Department of Energy and Climate Change suggested cutting feed-in tariffs — the price that an electricity utility must pay to generators

of solar energy — because too many large-scale projects were being planned. Under the revised system, subsidies for large arrays of photovoltaic panels would be slashed by 50–75% from 1 August.

## RESEARCH

## Arthropod project

A consortium of US and European scientists hopes to sequence the complete genomes of 5,000 species of arthropod within 5 years. The i5K Arthropod Sequencing Initiative will focus on economically important insects, disease vectors, model organisms and representatives from all the evolutionary branches of insects and related species. Currently, fewer than 50 arthropod genomes have been sequenced. The initiative was formally announced last week (G. E. Robinson *et al.* *Science* 331, 1386; 2011), and species can be nominated for sequencing at [www.arthropodgenomes.org/wiki/i5K](http://www.arthropodgenomes.org/wiki/i5K).

## Penguins in peril

Conservationists are warning of an ecological disaster in the south Atlantic after the cargo vessel *MS Oliva* broke up near Nightingale Island on 18 March, spilling heavy crude oil and threatening penguins already classed as endangered. The island



forms part of the Tristan da Cunha group, which is UK territory; tens of thousands of Northern Rockhopper penguins (*Eudyptes moseleyi*) live there. A slick from leaking oil extends 13 kilometres offshore, and oiled penguins (pictured) have been seen on the shores. See [go.nature.com/z11rlp](http://go.nature.com/z11rlp) for more.

## Alone in space

The next large science mission for the European Space Agency will have to make do without funding from NASA, the agency has decided. The 'L-class' mission, to launch around 2020, will be selected in February 2012 from three competitors, and will see its budget capped at €700 million (US\$996 million), about half of what had been hoped for. See page 421 for more.

## Orbiting Mercury

NASA's MESSENGER spacecraft has become the first probe ever to orbit Mercury. The craft slipped into orbit around the Solar System's innermost planet on 18 March,

after a 6.5-year flight that included three earlier flybys of its target. MESSENGER will conduct a comprehensive one-year survey of Mercury, studying surface features as small as 18 metres across, and searching for hints of ice within permanently shadowed craters near the planet's poles. It will also make magnetic-field measurements that could reveal structural details about Mercury's iron core. NASA expects the probe to start beaming imaging data back to Earth from 4 April.

## Crop genetics

An international treaty, whose 127 signatories pledge to share genetic information about food crops, has secured more than US\$10 million in donations for a second round of research grants, aimed at conserving global food security. Grant winners will be announced in May. Best known for its role in enabling the construction of the Svalbard Global Seed Vault in Norway, the treaty had struggled to gather research funds. The latest donations were announced at a meeting last week in Bali, Indonesia. See [go.nature.com/wyygdg](http://go.nature.com/wyygdg) for more.

## Fire above lab

A major underground laboratory at the bottom of the Soudan Mine in Minnesota was not seriously harmed by a

## COMING UP

### 27–31 MARCH

The American Chemical Society holds its spring meeting in Anaheim, California, focusing on the chemistry of natural resources.

[go.nature.com/egqhue](http://go.nature.com/egqhue)

### 28 MARCH

The Royal Society in London releases a report mapping how science is done around the world, and how these patterns are changing.

[go.nature.com/iv3bqu](http://go.nature.com/iv3bqu)

fire that broke out on 17 March in a lift shaft serving the lab. However, physics experiments there were suspended this week as crews worked to restart ventilation systems and pumps after quenching the fire. Among the high-energy experiments hosted by the University of Minnesota lab is the Cryogenic Dark Matter Search, which looks for signals of dark matter passing through crystals of germanium and silicon. See [go.nature.com/ma2spj](http://go.nature.com/ma2spj) for more and page 433 for a News Feature about dark-matter detectors.

## Clinical-trial access

The European Medicines Agency has started to allow public access to its database of clinical trials. Information on interventional trials that are being carried out in all 27 European Union (EU) member states, as well as in Iceland, Liechtenstein and Norway, will be searchable in the EU Clinical Trials Register ([go.nature.com/cacwil](http://go.nature.com/cacwil)). As part of transparency measures, the site will gradually publish information from EudraCT, the EU's database of clinical trials dating from May 2004, which is not publicly available at the moment.

**NATURE.COM**

For daily news updates see:

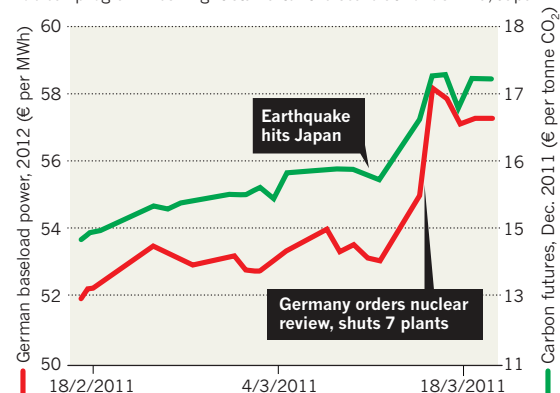
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

Jumps in the prices of electricity and European carbon permits accompanied the German government's 15 March decision to shut the country's seven oldest nuclear reactors, and to suspend the planned extension of licences to operate other nuclear plants. To compensate for the drop in nuclear energy supply, utilities companies expect to switch to producing electricity from more carbon-intensive gas and coal, thus boosting demand for future allowances to emit carbon dioxide on the Emissions Trading System.

## NUCLEAR CRISIS RATTLES EUROPEAN ENERGY MARKETS

Electricity and carbon-permit prices rose as investors worried that civil nuclear programmes might stall after the scare at Fukushima, Japan.

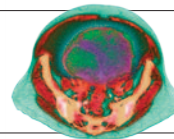


# NEWS IN FOCUS

**JAPAN** How the earthquake shook up the nation's science infrastructure **p.420**

**AUSTRALIA** A heated debate breaks out over cattle and bushfire **p.422**

**GENOMICS** What is the best way to assemble a genome? **p.425**



**CANCER DIAGNOSIS** The troubled search for a biomarker **p.428**

TOKYO ELECTRIC POWER CO./PRESS ASSOCIATION IMAGES



Smoke rising from a reactor at the Fukushima Daiichi nuclear plant meant workers had to be evacuated.

JAPAN

## The meltdown that wasn't

*How a handful of operators at a crippled reactor averted a greater catastrophe at the Fukushima plant.*

BY GEOFF BRUMFIEL

The magnitude-9.0 earthquake rocked Fukushima Daiichi nuclear power station at 2:46 p.m. on 11 March, but the real emergency began an hour later. A wall of water swept across the site, washing away power lines and the fuel tanks for the emergency backup generators designed to take over

if grid power failed. Inside the control room of the unit 1 reactor, the lights went out and the 1970s-vintage analogue gauges drifted to zero.

It will probably be years before anyone knows exactly what happened inside the three reactors at Fukushima Daiichi that seem to have partially melted down in the wake of the tsunami. But from press reports, public statements and interviews with experts, it is

possible to work out the most likely scenario. And already it is clear that decisions made in the initial 24 hours by the handful of operators in the control room probably averted a much greater nuclear catastrophe than the one that now faces Japan.

In the moments after the power was lost, the operators “would have literally been blind”, says Margaret Harding, a nuclear engineer in Wilmington, North Carolina. Harding worked for two decades with General Electric, which designed Fukushima's boiling-water reactors, and she witnessed a similar outage in 1984 during a safety test at a boiling-water reactor in Switzerland. “Basically the emergency lights came on and all the panels went black,” Harding says.

During the Swiss test, the power returned in 5 minutes. At Fukushima, batteries ran a handful of emergency lights in the control room and a few instruments tracking the reactor's vital signs, such as the pressure inside the core.

The core was next door. Inside a large, cube-shaped building, enclosed in a heavy concrete containment vessel, sat a thick, steel capsule filled with around 50 tonnes of uranium. Until an hour previously, that fuel had been pumping out 460 megawatts of power, but the reactor had automatically shut down immediately after the earthquake. Boron-carbon control rods driven between the long columns of fuel had soaked up neutrons and halted the nuclear reactions.

### MODEL RESPONSE

That didn't mean the reactor was cold. Radioactive by-products of the fission reactions still generated heat — some 7 megawatts of it, preliminary computer models by the National Nuclear Laboratory in Sellafield, UK, suggest. The fuel still needed to be actively cooled.

Without power, operators could use steam from the reactor's pressure vessel, plus minimal amounts of battery power, to drive a pump that would keep the cooling water circulating. What they probably didn't know was that the cooling system had sprung a leak. The leak caused water levels inside the core to drop, allowing the fuel to heat up, which generated more steam and raised the pressure inside the steel vessel. The emergency cooling system was unable to cope, according to a press release from the Tokyo Electric Power Company (TEPCO), the plant's

**NATURE.COM**  
For full coverage of the earthquake and nuclear crisis, see: [go.nature.com/ulsz2n](http://go.nature.com/ulsz2n)

emergency cooling system was unable to cope, according to a press release from the Tokyo Electric Power Company (TEPCO), the plant's



## AN UNFOLDING CRISIS

At the Fukushima Daiichi power station, the plant hit hardest by the earthquake and tsunami, unit 1 was the first of four operating reactors to reach a point of crisis.

### FRI 11 MARCH

**14:46** The M9.0 earthquake struck; reactors were shut down.

**15:42** Mains power was lost.

**15:45** Oil tanks were washed away by tsunami.

**16:36** Emergency core cooling system failed in **unit 1** and **unit 2**.

**19:03** A nuclear emergency was declared at Fukushima Daiichi.

**20:50** Residents living within 2 km of Fukushima Daiichi were told to evacuate.

### SAT 12 MARCH

**4:00** Pressure in the containment at **unit 1** reaches 840 kPa, twice the design value.

**5:44** Evacuation was expanded to a 10-km radius of Fukushima Daiichi.

**14:30** Steam was vented from **unit 1**.

**14:49** Radioactive caesium was detected around **unit 1**.

**15:36** Hydrogen explosion took place at **unit 1**.

**20:20** Sea water was injected into the reactor at **unit 1**.

### SUN 13 MARCH

**13:12** Sea water was injected into the reactor at **unit 3**.

### MON 14 MARCH

**4:08** Spent fuel pool at **unit 4** overheated.

**11:01** **Unit 3** exploded.

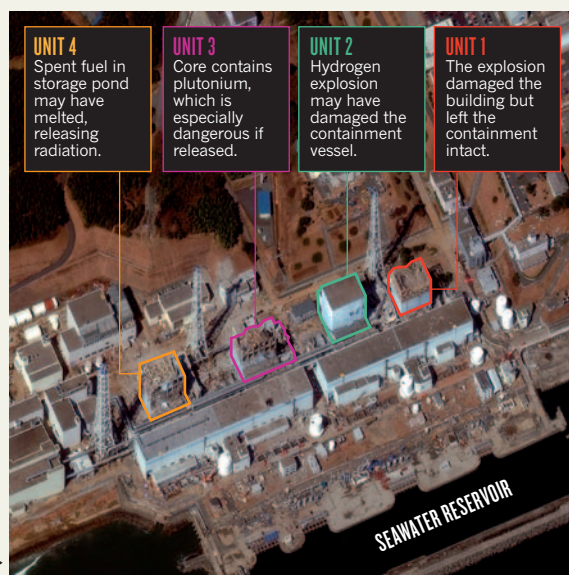
**16:34** Sea water was injected into the reactor at **unit 2**.

### TUE 15 MARCH

**6:10** **Unit 2** exploded.

**9:38** Fire broke out at the reactor building of **unit 4**.

A satellite view of the Fukushima Daiichi plant shows heavy damage from explosions or fire at three of the four affected reactors.



DIGITALGLOBE/REUTERS

► operator. At 7:03 p.m. a state of nuclear emergency was declared. Less than 2 hours later, evacuations began within a 2-kilometre radius of the plant.

By 4:00 a.m. the pressure inside the thick steel vessel of unit 1 had reached 840 kilopascals (kPa), more than twice the operating limit, according to the Nuclear and Industrial Safety Agency, the Japanese nuclear regulator. Radiation levels at the front gate of the site had begun to rise above background, although they were still far from dangerous. At 5:44 a.m. the evacuation cordon was expanded to 10 kilometres.

At some point, the falling water levels must have left the fuel exposed. In a reactor such as unit 1, the uranium pellets are enclosed in long, skinny pipes made of zirconium alloy, chosen because it does not inhibit the neutrons needed to drive the fission reactions. As temperatures rose above 1,000 °C, the steam in the pressure vessel began to oxidize the zirconium, probably releasing hydrogen gas. Meanwhile, fuel pellets, liberated from their shell, began to fall to the bottom of the reactor. The meltdown had begun.

This was the crucial moment. If the operators at unit 1 could not stem the meltdown, the fuel would gather at the bottom of the vessel. The uranium pellets, now in close proximity, could begin exchanging neutrons and resume their heat-producing nuclear reactions. Slowly, the pile could build towards a 'critical mass' that would restart the nuclear process normally used to generate electricity.

### TURNING POINT

Nobody can be sure about this sequence of events because there has never been a full meltdown in a boiling-water reactor. Harding says that she thinks it's unlikely that the nuclear processes would have reignited. Even if they did, the worst case, in her opinion, is

that the fuel would have burned through the steel pressure vessel and splattered onto the 'base mat', a thick concrete slab that would have spread out the fuel, extinguishing any fission reactions.

But even that might have been catastrophic. The volatile hydrogen gas generated by the zirconium was safe inside the steel pressure vessel, but it was liable to explode if exposed to air in the outer containment vessel. If the blast were big enough, it might have breached the outer vessel's thick, concrete walls.

This scenario is highly unlikely, but had it happened, the workers struggling to save the plant would almost certainly have received a

***"The fire trucks were brilliant. I'm not sure I would have thought of that."***

lethal dose of radiation, says Malcolm Sperrin, a medical physicist at the Royal Berkshire Hospital in Reading, UK. Citizens near the plant could have been at higher risk for cancer later in life, he says. And the contamination would have made emergency operations much more difficult at the other reactors, which were also in trouble. The situation could easily have spiralled out of control.

Just metres away was a vast reservoir of sea water. It could stop the reactor's meltdown, but operators had no way to pump it into the core. Emergency generators could not be hooked into the system, for reasons that are still unclear.

At some point, somebody on the site realized that fire engines were essentially giant portable pumps with their own power supplies. "The fire trucks were brilliant," Harding says, "I'm not sure I would have thought of that." Engines were rushed to the plant and hooked into the lifeless emergency cooling system. Yet there was still a problem: the pressure in the core was too high for the engines' pumps to force in the sea water.

Around 2:30 on Saturday afternoon, operators began to vent pressure from the containment vessel. An hour later something sparked the gas that had built up inside the outer building during venting. The entire top of unit 1 was blown away, and four workers were injured, although the sturdy concrete containment vessel below seems to have survived the blast.

### CHAIN REACTION

The explosion, broadcast around the world, was the first of a series of setbacks at the reactor complex. In the ensuing days, reactors 3 and 2 followed a similar path to unit 1 (see 'An unfolding crisis'); each was rocked by a massive hydrogen explosion. In units 3 and 4, the pools for storing used fuel lost their cooling water and it is believed that the rods began to melt, emitting more explosive hydrogen along with powerful radiation.

At the time of writing, radioactive material from Fukushima Daiichi continues to blow across Japan at levels high enough to cause concern for Sperrin — although he says that they are not immediately dangerous. In the coming weeks and months, the government, TEPCO and safety authorities are likely to face heavy criticism. People will ask what went wrong.

Still, at unit 1 the immediate crisis has passed. With the pressure down, fire engines began to flood the reactor with sea water at 8:20 p.m. on 12 March, allowing the fuel to slowly cool to a safe temperature. The response at unit 1 also provided a model for stabilizing the other two reactors. And day by day, the radioactive decay in the reactor cores is ebbing. It could be days or weeks before the reactors are truly safe, but for now things remain stable.

As for the operators at unit 1, says Harding, "I think they really did respond pretty well." ■

JAPAN

# Radiation risks unknown

*Scientists struggle to calculate long-term effects of low-dose exposures in Fukushima.*

BY GWYNETH DICKEY ZAKAIB

One thing is certain about the human costs of the radiation leaking from the Fukushima Daiichi nuclear plant in Japan: they will pale in comparison to the catastrophic consequences of the 11 March earthquake and tsunami that triggered the crisis. Nevertheless, experts are tracking radiation levels worldwide to learn more about the accident and to assess the possible impacts on health.

Radioactive vapour and particles released from the plant have spread across the region and followed prevailing winds across the Pacific (see 'Plume projections'). "The plume is very large," says Ted Bowyer, a nuclear physicist at the Pacific Northwest National Laboratory in Richland, Washington, one of the first US stations to detect isotopes released from Fukushima. Bowyer adds that the tiny concentrations of radioactive iodine, caesium, tellurium, xenon and lanthanum that have reached the United States are far below normal background levels and not a health risk. The fact that some of the isotopes are short-lived indicates that at least some of the radiation must have originated from breaches in the reactor vessels and not from the plant's overheated caches of spent fuel, he says.



A farmer destroys spinach in Ibaraki prefecture after it was contaminated by radioactive iodine.

In Fukushima and adjacent prefectures, the Japanese government is reporting radioactive contamination in sea water near the plant and in the food and water supply. Radioactive iodine-131 and caesium-137 have been detected in milk and leafy vegetables such as spinach, as well as in tap water, in some cases above allowable levels for consumption. Such safety limits are based on long-term consumption of these foods, says William McCarthy, deputy director of the radiation protection programme within the Environment, Health and Safety Office at the Massachusetts Institute of Technology (MIT) in Cambridge. "The prudent thing is to not eat that food," he says. "That doesn't mean it poses immediate health risks."

Authorities in Japan have banned the shipment of milk from Fukushima prefecture, as well as some produce from Fukushima and three neighbouring prefectures. In the short term, the main concern is iodine-131, which can cause cancer in the thyroid gland. With a half life of 8 days, iodine-131 will effectively be gone from the environment in a matter of months once releases have stopped. But

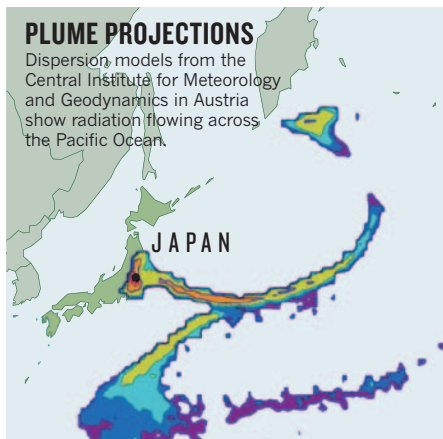
caesium-137, another cancer-causing isotope, has a half-life of 30 years and will persist for much longer. Steve Wing, an epidemiologist from the University of North Carolina, Chapel Hill, points out that even the low levels of radiation that remain in the environment could be significant in the long run "because so many more people are exposed, even though the dose per person decreases farther from the plant".

Jacquelyn Yanch, a radiation physicist at MIT, thinks that it is too early to say what the impact will be. "We haven't come up with risk estimates for a situation like this," she says. "We don't know how much is too much."

Experts agree that any long-term effects are most likely to be seen in the workers battling the crisis at the Fukushima nuclear station. The government has increased the allowable dose for workers from 100 millisieverts per year to 250 millisieverts per year — five times the annual allowable dose for US radiation workers — to allow emergency operations to continue. This dose is considered by the US National Institutes of Health as the lower limit for the first symptoms of radiation sickness. ■

## PLUME PROJECTIONS

Dispersion models from the Central Institute for Meteorology and Geodynamics in Austria show radiation flowing across the Pacific Ocean.



SOURCE: CENTRAL INST. FOR METEOROLOGY AND GEODYNAMICS



**MORE  
ONLINE**

### TOP NEWS



Journal takes another look at controversial home-birth study [go.nature.com/py58xr](http://go.nature.com/py58xr)

### MORE NEWS

- Eyes in the sky track Libyan abuses [go.nature.com/qbhdz4](http://go.nature.com/qbhdz4)
- China plans science-funding reform [go.nature.com/lshyys](http://go.nature.com/lshyys)
- Mining plans cause problems for South African astronomers [go.nature.com/liiy2f](http://go.nature.com/liiy2f)

### ON THE BLOG



United States strikes deal with environmentalists to allow culling of wolves to go ahead [go.nature.com/sil5we](http://go.nature.com/sil5we)

OCEAN/CORBIS

USFWS





Some buildings at Tohoku University are too damaged to be entered safely.

JAPAN

# Quake shakes Japan's science

*Natural disaster leaves researchers struggling with broken equipment and a crippled infrastructure.*

BY ICHIKO FUYUNO

The magnitude-9.0 earthquake that struck northeastern Japan on 11 March trashed Koji Tamura's laboratory and office, flinging books, microscopes, sequencers and samples to the floor. The geckos, *Xenopus* frogs and zebrafish that the Tohoku University researcher uses to study organ development survived the quake but now face a slow death, because disrupted water supplies mean their tanks may run dry. "Without water, I am worried how long our animals can survive," he says. "But I cannot think about research at a time when many suffering people need water to live."

Like many scientists in Japan, Tamura is both anxious over disrupted research plans and heartbroken at the human toll of the earthquake and the ensuing tsunami, which killed thousands of people and left nearly half a million homeless. No casualties have been reported on campus at universities and research institutes so far. But more than a week after the event, scientists taking stock of the damage foresee a long, difficult recovery, as disrupted infrastructure and power outages in

Tokyo and other eastern parts of Japan add to the physical damage at labs and other facilities.

The earthquake hit hardest at Tohoku University, a materials-science, engineering and biomedicine powerhouse in the city of Sendai, close to the epicentre. At the university, which is expected to remain closed until late April, an emergency team is assessing the damage, but scarce electricity, gas and water, coupled with intermittent aftershocks, are making inspections extremely difficult. The tsunami flooded one building at a field station of the Graduate School of Agricultural Science on the coast north of Sendai, and six buildings on the main campus are too dangerous to enter.

The university's WPI-Advanced Institute for Materials Research, renowned for its work on metallic glasses, polymers and nanodevices, has lost ¥1 billion (US\$12.5 million) of equipment, and the cost is likely to increase when the damage is assessed in detail, says Yoshinori Yamamoto, the institute's director. Broken instruments include some of the world's best electron microscopes and instruments for studying the atomic arrangement of surfaces.

Farther from the epicentre, the Japan Proton Accelerator Research Complex (J-PARC),

south of Sendai on the coastline of Ibaraki prefecture, has been shut down. Its three accelerators seem to be intact, and the facility escaped damage from the tsunami. But the earthquake cut off the water supply, buckled nearby roads and damaged computer servers. This week researchers will begin switching on the facilities for preliminary inspections.

The High-Energy Accelerator Research Organization in Tsukuba has also seen its Photon Factory synchrotron crippled. "The linear accelerator has seen some substantial damages," Soichi Wakatsuki, director of the Photon Factory, wrote to the international community on 15 March. But he noted that "five protein crystallography beamlines have been spared of major damages".

Meanwhile, the shutdowns of the Fukushima Daiichi reactor complex and other nuclear plants after the earthquake have led to electricity shortages in Tokyo and neighbouring prefectures, where temporary outages are planned every day. Many institutions in the region, including the University of Tokyo and some RIKEN institutes, have been forced to drastically reduce electricity use and shut down large facilities such as supercomputers.

Uncertainty resulting from the disrupted infrastructure and the nuclear crisis is prompting foreign nationals to decamp for cities farther south or overseas. "That is the major problem," says Adrian Moore, a unit leader of the Brain Science Institute at RIKEN in Wako. Five out of his six non-Japanese postdocs and students have left and will not return until the situation improves. At J-PARC, all foreign researchers have flown home or are being housed in dormitories in Tsukuba, where the infrastructure held up better. The government is considering emergency funding for rebuilding universities and research institutes.

Amid the depressing circumstances, there are bright notes. On the day of the earthquake, Japan's research vessel *Chikyu*, capable of drilling seven kilometres into the sea floor, was docked in Hachinohe, north of Sendai, preparing for a voyage to sample coal beds deep under the sea floor.

Within 30 minutes of the quake, the ship undocked with 200 people aboard, including 48 elementary school pupils on a tour, because a ship at sea is safer than the shore during a tsunami. The tsunami arrived almost immediately afterwards, spinning the 57,000-tonne vessel 2.5 times but causing no injuries. The only damage was to one of the ship's six thrusters. "It was almost a miracle that no life was lost," says Fumio Inagaki, co-chief scientist of the coal-bed expedition at the Japan Agency for Marine-Earth Science and Technology. The upcoming voyage was cancelled, however.

Yukihisa Kitamura, executive vice-president of Tohoku University, says students and faculty members are encouraged by support messages from around the world. "We are regaining our enthusiasm rather than giving in," he says. ■



## SPACE SCIENCE

# Europe makes do without NASA

*US budget crisis forces European Space Agency to abandon plans for joint mission.*

BY EUGENIE SAMUEL REICH

**T**he European Space Agency (ESA) is pushing ahead without NASA support for its next big science mission, as the ongoing US budget crunch and competing priorities impose serious constraints on the US space agency (see *Nature* 471, 278; 2011). ESA last week told leaders of three large, or 'L-class', missions that are competing for funding to revise their proposals by leaving out the substantial US contribution that had previously been assumed.

"The decision was made very reluctantly," says David Southwood, director of science and robotic exploration at ESA. "NASA could not meet our timetable to launch."

The budget available to the winning L-class mission, set to launch in 2020, will be cut by 40–50% to a round figure of €700 million (US\$996 million). The competing missions are the International X-ray Observatory (IXO), a telescope that would observe black holes and

the formation of galaxies; the Laser Interferometer Space Antenna (LISA), which would detect gravitational waves; and the Europa Jupiter System Mission (EJSM-Laplace), in which twin probes would head to Jupiter's icy moons Europa and Ganymede. ESA had planned to choose a winner in June, but it has now put off the decision until February 2012 while teams rework their proposals.

The reduced budget will affect the science capabilities of each mission, but team leaders are optimistic. Joel Bregman, a member of the science definition team for IXO, says that even if the telescope is built with a smaller mirror than planned and only two or three instruments instead of the proposed six, it will still be a step ahead of current X-ray missions. And Karsten Danzmann, European co-chairman of the LISA international science team, says that a more modest LISA could still detect gravitational

waves, but would be sensitive to waves emanating from fewer kinds of astrophysical source.

Ronald Greeley, US co-chairman of the EJSM-Laplace science team, says that his group's mission may be the easiest to rework, because it involves two spacecraft. The NASA-led Europa probe could be dropped — although the mission would lose much of its scientific appeal.

NASA still hopes to contribute in smaller ways to ESA's L-class mission, but it faces shortfalls with a flagship programme of its own: James Green, director of NASA's planetary-science division, told an advisory committee on 16 March that under President Barack Obama's 2012 budget request, the agency has about \$1 billion to commit to phase one of its top planetary-science priority, a sample-return mission to Mars. The mission is expected to cost far more, so it will not be able to go forward without a comparable contribution from ESA. A bilateral meeting between NASA and ESA officials is set for 29–30 March in Pasadena, California. ■

**NATURE.COM**

For more on ESA's proposed L-class missions, visit: [go.nature.com/ilu2m9](http://go.nature.com/ilu2m9)



Cattle have been released into parts of Australia's Alpine National Park, despite a ban on grazing there.

# ECOLOGY

## Australian grazing trial ignites debate

*Researchers question science behind controversial effort to examine cattle's role in controlling bushfires.*

BY EMMA MARRIS

To ecologists overseas, the invitation might sound tempting. It offers travel to Australia and unspecified remuneration to serve on an advisory panel considering a juicy scientific question: could allowing cattle to graze in the country's Alpine National Park — the picturesque setting for the film *The Man from Snowy River* — reduce the risk of bushfires? Those responding to the call, issued by the federal state of Victoria at the beginning of March, might not realize that it also involves walking into a serious stoush — that's Australian for fight.

More than 100 Australian ecologists have signed a letter to the Australian government denouncing the trial. The letter's organizers claim that the trial is a naked attempt to use the imprimatur of science to allow cattle to graze in an ecologically sensitive area, and they now fear that international scientists ignorant of the ruse will be duped into lending credibility to the project. "It's a misuse of the word science to justify a political decision," says Georgia Garrard, an ecologist at the University of Melbourne who helped to organize the protest letter.

The decision in question is the return of cattle to portions of the 646,000-hectare park, a landscape of deep ravines, high plateaux and snow gum trees. The move fulfils a campaign

pledge by Victoria's centre-right coalition, which was narrowly elected last November.

In the 1950s, as many as 100,000 head of Aberdeen Angus and Hereford grazed in the park, according to Mark Coleman, president of the Mountain Cattlemen's Association of Victoria. Gradual restrictions since then culminated in a park-wide grazing ban in 2005 after a Victoria government review found that cattle didn't reduce the intense wildfires that can visit the region, but that their hooves, grazing and manure damage sensitive wetland ecosystems.

Coleman says that the cattle do stop fires, by eating the vegetation that forms potential fuel, and that ecologists have ignored mountain cattlemen's knowledge of the land. "We've got generational knowledge that goes back — 150 years in my family. You can't buy that knowledge; you can't learn it in a university."

In January, when 400 head of cattle were allowed back into the park, the government of Victoria announced that they were part of a research trial. Opponents of the move argue that no baseline data were taken and that the trial has not been designed. Designing the trial will be one of the advisory panel's tasks, says Peter Appleford, executive director of

forests and parks for Victoria. He defends the lack of baseline data, saying that, for a large landscape-scale trial, such data would need to have been kept for decades to be relevant. "Some of these scientists have been trained on small plots and don't understand landscape-scale experiments," he adds.

Libby Rumpff, an ecologist at the University of Melbourne and an organizer of the scientific protest, disagrees. "It doesn't matter what scale you are looking at, you should be collecting baseline data," she says.

The Melbourne newspaper *The Sunday Age* has charged the Victoria government with trying to blackmail the University of Melbourne into participating in the trial by threatening its state funding, a charge the government denies. And John DuBois, a spokesman for the university, says that researchers at its School of Land and Environment have told the government that the trial may amount to repetition of previous work. The university has no plans to participate directly.

Mark Adams, an ecologist at the University of Sydney, has been invited by the Victoria government to "lead the program of research that will accompany" the trial, according to Victoria's website. Adams says that no contract has been signed between the university and the government, "and without a contract I won't be doing anything." If a contract is signed, he adds, he won't be designing a trial, but merely "testing methods suitable for measuring the impact of cattle on fuels and on ecosystem functions". Adams says that he has taken a lot of flak for his willingness to work with Victoria, but feels that academics "have a duty to work with government to try to get the best outcomes in the public interest".

On 18 March, Tony Burke, Australia's environment minister, demanded that the state ask the federal government for approval to release the cattle. The demand means that the cattle must be removed from

the park until the federal government has reviewed the matter. But even without this demand, the animals' time in the highlands is almost up. Winter snows will arrive soon, and the cattle would have been taken down from the mountains by mid-April anyway. Furthermore, it is

not clear what the federal decision means for the "excellent international candidates" who Appleford says have already applied to sit on the advisory panel.

The point of the episode, says Garrard, is "that science shouldn't be misused for political gain". Coleman sees both sides as politicized. The ecologists have been "indoctrinated" by politicians on the left, he says, whereas the coalition is trying to deliver on a political promise. "Ninety per cent of the general public couldn't give a shit about the cattle or the environment," he says. "At the end of the day, it is all politics." ■



## COMMUNITY

# US government scientists test limits of conflict rules

*Policy change has made it easier to serve on boards of scientific societies.*

BY EUGENIE SAMUEL REICH

When Mike McPhaden was elected president of the American Geophysical Union (AGU) last year, he was delighted — but he wasn't sure he would be able to take up the position. McPhaden is an oceanographer at the Pacific Marine Environmental Laboratory in Seattle, Washington, which is run by the National Oceanic and Atmospheric Administration (NOAA). Lawyers at the Department of Commerce, which oversees NOAA, were concerned that leading a scientific organization that lobbies the government on funding and policy matters would create a conflict of interest for McPhaden. "There was resistance," he says.

In the end, McPhaden convinced the agency that taking up the position would bring prestige to his government role and enhance the credibility of NOAA science. Today, a memorandum of understanding between the AGU and NOAA even allows him to spend some of his government-paid time working for the scientific society, although he has to recuse himself from both fund-raising and lobbying.

Now, changes in US government policy should make it much easier for government scientists to serve in scientific societies. A memorandum on scientific integrity issued by the Office of Science and Technology Policy (OSTP) in December explicitly encourages government scientists to get involved with societies; previously, the government tended to view such associations ambivalently or negatively. Yet many government scientists affected by the policy change say that serious legal and ethical pitfalls remain.

Unlike in countries such as Britain, which has no rule against government scientists serving on society boards, the strict conflict-of-interest rules in the United States can create administrative barriers for government scientists trying to participate in societies that are relevant to their disciplines. Under US law, for example, government officials are barred from participating in

matters in which they or organizations they are associated with have a financial interest. In some cases, the restrictions have been interpreted as preventing government employees from lobbying. Employees who join outside organizations will have to be careful not to run afoul of these rules, notes John Fitzgerald, policy director of the Society for Conservation Biology in Washington DC. Fitzgerald supports the more permissive policy but cautions that a government scientist who lobbies Congress could be "skating on thin ice".

A scientist who has navigated that issue

turned down the offer. The Fish and Wildlife Service gives the society a few thousand dollars each year to spend on scientific meetings. She was worried that some might think she got the prestigious board position in exchange for ensuring that the funding continued, and her colleagues shared her concerns. "At the end of the day it's about credibility," she says.

## DIFFERENT EMPHASIS

The US Geological Survey (USGS) has a more liberal approach. USGS scientists are encouraged to serve in scientific societies, and 91 currently do so. Their promotional prospects depend on them showing leadership in the research community — which they can do by being elected to a society board. However, USGS director Marcia McNutt says that conflicts of interest are less likely to arise at the USGS because the agency has no policy-making authority.

Both the Fish and Wildlife Service and the USGS are run by the US Department of the Interior (DOI), and McNutt is pleased that the department has now consolidated its scientific-integrity policy. The new policy, introduced on 1 February, follows the OSTP guidance and explicitly encourages all researchers within

the DOI's jurisdiction to participate in scientific societies, although they need to fill out forms before going ahead. That paperwork helps to ensure that researchers understand what kinds of behaviour could be considered a conflict — for example, serving on the board of a society and then signing a government purchase order for the society's publications. "Scientists can be clueless about the trouble they can get themselves into," says McNutt.

Various other agencies are now expected to work the OSTP guidance into their policies, and watchdog groups are delighted at the changing attitude. Jeff Ruch, executive director of Public Employees for Environmental Responsibility in Washington DC, says that government lawyers have long been allowed to participate in professional organizations such as the American Bar Association, and says he doesn't see why the practice shouldn't be extended to scientists. ■



William Talman takes leave from his government job to lobby Congress.

is William Talman, president of the Federation of American Societies for Experimental Biology in Washington DC, who also works as a physician at a hospital in Iowa City run by the US government's Department of Veterans Affairs. Acting as president of the federation, Talman has written to senators to advocate for generous funding for the National Institutes of Health and testified in Congress in support of budget boosts for the US National Science Foundation. Talman says that he takes unpaid leave from his government job for those activities, an arrangement that happens to fit with the new policy. As a result, he says, "I would argue there's no conflict of interest."

Biologist Gabriela Chavarria, who is science adviser to the director of the US Fish and Wildlife Service, worried about a different issue when the Society for Conservation Biology invited her to serve on its board earlier this month. Following discussions with the agency, Chavarria

➔ **NATURE.COM**  
For more on the  
OSTP guidance, see  
[go.nature.com/ter5uj](http://go.nature.com/ter5uj)



## GENOMICS

# Genome builders face the competition

*Three independent projects seek to contrast approaches in preparation for routine analysis of genetic data.*

BY ERIKA CHECK HAYDEN

Sequencing DNA on an industrial scale is no longer difficult: the challenge is in assembling a full genome from the multitude of short, overlapping snippets that second-generation sequencing machines churn out. Researchers can call on any of two dozen computer programs to do the job, but all have their flaws. With genome sequencing fast becoming standard practice across the life sciences, researchers want to know how to choose.

The answer may come from three separate genome-assembly projects, each of which aims to test different algorithms on batches of raw sequence data and to compare the results.

There won't be any single 'winner', researchers stress. There is no consensus way to determine the absolute quality of a genome, and different assemblers might do a better job of handling different types of data.

"My dream is that a few years from now, a person who is about to do a genome project will be able to say, 'This is our budget, these are the characteristics of our genome; what is the combination of sequencing technologies and genome-assembly program that best fits our project?'" says Ian Korf at the University of California, Davis, who helped to organize the Assemblathon, one of the three genome-assembly evaluation projects.

Last December, the Assemblathon released a computer-generated human genome data set. Scientists were invited to use their assembler of choice to stitch the data into a genome. Seventeen teams from seven countries took up the challenge. Korf's team then evaluated the assemblies on the basis of commonly used criteria for the quality of genome assemblies — such as the portion of the genome that is assembled into large chunks of DNA, or contigs — as well as less-common measurements, such as how many genes each assembly is able to capture.

At a meeting last week at the University of California, Santa Cruz, three winners emerged: ALLPATHS-LG, developed by the Broad

Institute in Cambridge, Massachusetts; ABySS, developed at the British Columbia Cancer Agency's Genome Sciences Centre in Vancouver, Canada; and SOAPdenovo, developed by the Beijing Genomics Institute. But, Korf notes, "it's not just the software, it's how people are running it" that determines the quality of each assembly.

A similar genome-assembly project called dnGASP has been organized by the National Center for Genome Analysis in Barcelona, Spain. Its results are set to be discussed at a workshop on 4–7 April.

A third project, led by Steven Salzberg of the University of Maryland, College Park, is evaluating just five assemblers, among them ALLPATHS-LG and SOAPdenovo. Salzberg's group will perform and evaluate all the assemblies. In

addition, the researchers will use real genome data from four species, including the Argentine ant and the common eastern bumblebee. "With purely simulated data, you don't get a realistic picture of how these assemblers perform," says Salzberg.

Later this year, the Assemblathon will launch another round of evaluation, comparing efforts to

assemble two previously unreleased genomes, that of a parrot and a cichlid fish. And although the three current efforts are focused on data generated by the popular Illumina sequencers, new sequencing methods could become commercially available as early as next year.

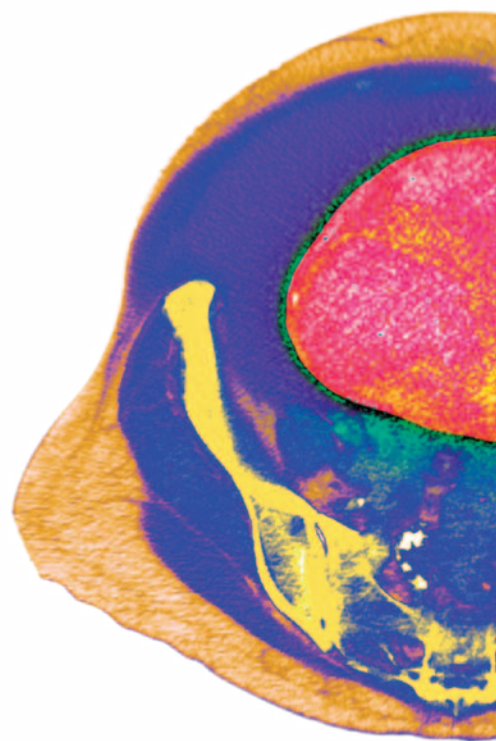
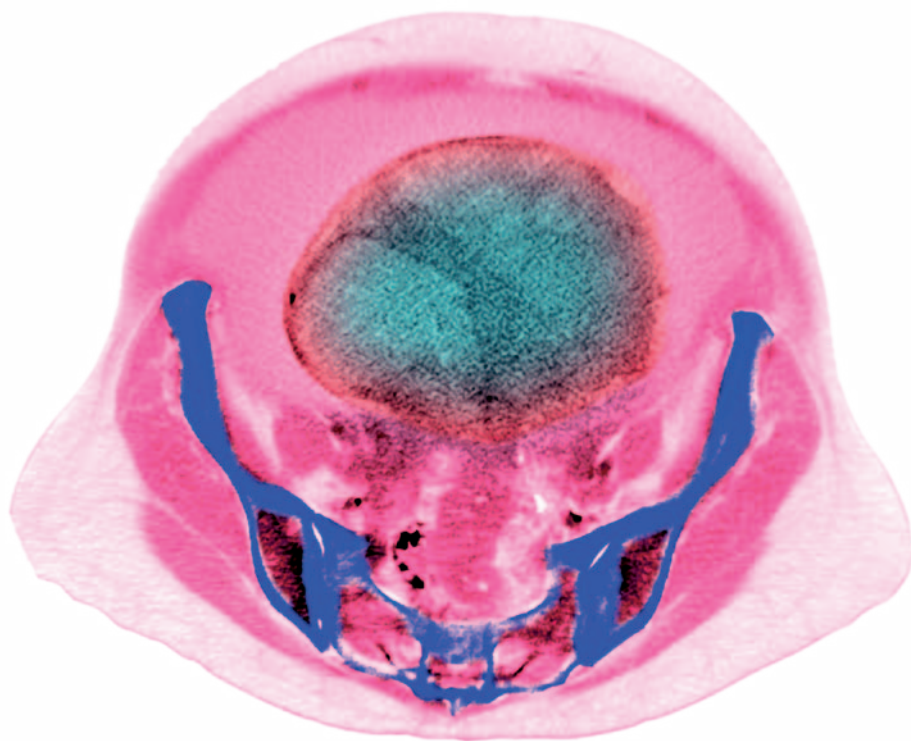
Their output will differ from that of the Illumina machines; the single molecule, real-time (SMRT) technology developed by Pacific Biosciences of Menlo Park, California, for instance, produces longer reads but has higher error rates (see *Nature* **470**, 155; 2011). This creates a new challenge, says Gene Robinson, an entomologist at the University of Illinois at Urbana-Champaign, whose bee sequence data are being used by the University of Maryland project. "Biologists really want assembly algorithms that can make use of multiple forms of reads and build the best possible assembly," Robinson says.

The contest is just beginning. ■



**Test case:** researchers will evaluate programs assembling this bee's genome.

D. WROBEL/VISUALS UNLIMITED/CORBIS



# MISSING THE MARK

*Why is it so hard to find a test to predict cancer?*

BY LIZZIE BUCHEN

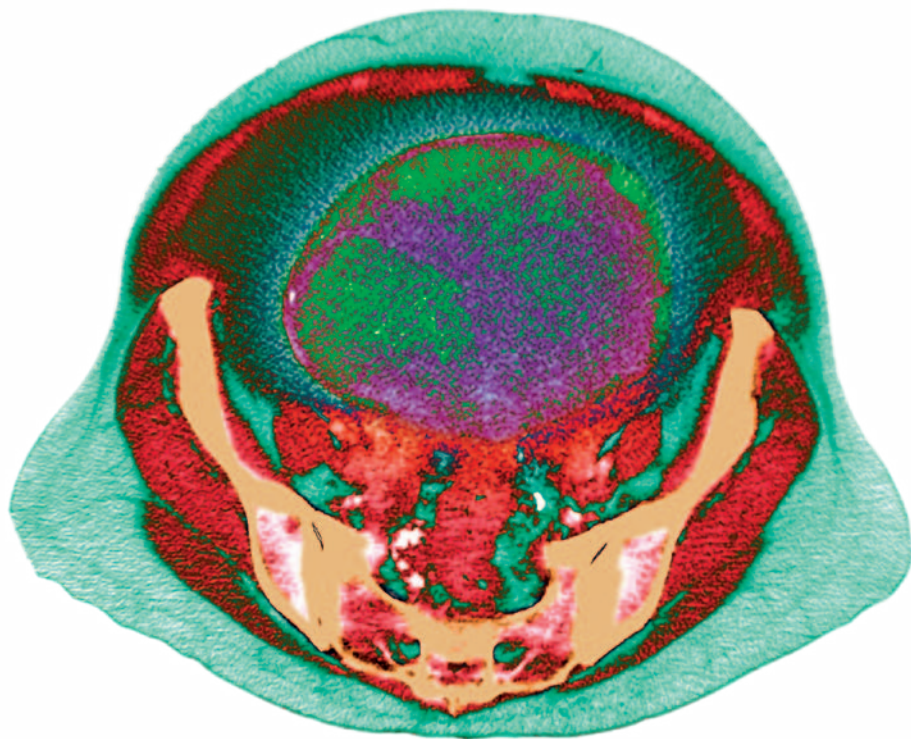
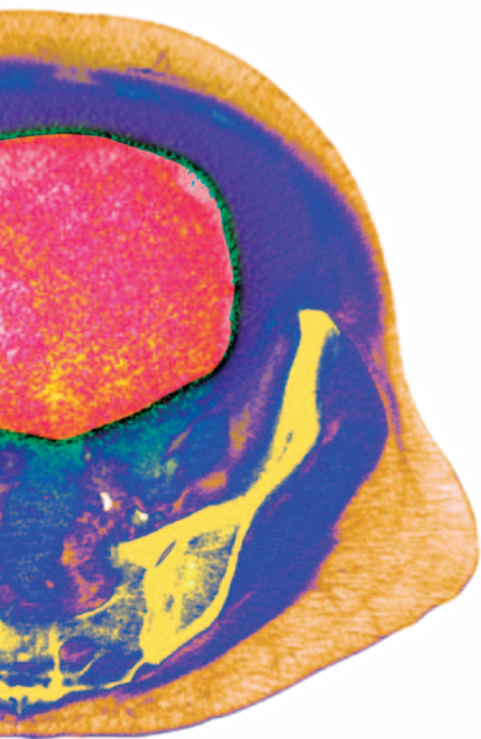
**O**n 3 March, two studies appeared online that offered 19 pages of gloomy reading for anyone interested in cancer. They focused on biological molecules, or biomarkers, the presence of which in the blood might be used to detect the earliest glimmers of ovarian cancer — a disease not normally discovered until it has destroyed the ovaries and rotted other parts of the body. The researchers, coordinated by the Early Detection Research Network (EDRN) of the US National Cancer Institute (NCI), had assembled 35 protein biomarkers, including 5 panels of proteins, that had looked the most promising in early studies. They had carried out rigorous testing — screening blood samples from more than 1,000

women — to ask whether these seemingly breakthrough biomarkers were better at identifying women with early ovarian cancer than the one flawed biomarker that had been in use for almost 30 years, CA-125. None of them was<sup>1,2</sup>. “CA-125 remains the ‘best of a bad lot,’” read an accompanying perspective article<sup>3</sup>. “The new candidates have fallen short of expectations.”

Tied in last place for its poor performance among the biomarker panels was one identified by Gil Mor, a cancer biologist at Yale University in New Haven, Connecticut. Mor’s six-protein panel detected ovarian cancer in only 34% of the women who were diagnosed with the disease within a year. (CA-125, by

contrast, detected 63%.) Mor’s panel already had a tortured history. A primary research paper behind it had been criticized by other scientists for allegedly using inappropriate statistical calculations and for optimistically concluding that the test would help women before rigorous follow-up studies proved that it could. Yet for four months in 2008, the test was sold to patients by Laboratory Corporation of America (LabCorp) in Burlington, North Carolina, the company that licensed the panel from Yale. LabCorp had marketed the test under the name OvaSure until the US Food and Drug Administration (FDA) intervened and the company pulled it from the market. The panel offered “invaluable object lessons”





**Biomarkers might help to detect ovarian tumours (large round masses) early in the disease.**

for bringing a test prematurely to the clinic, wrote the authors of the perspective article.

Similar lessons can be found in the stories behind many cancer biomarkers that have sputtered and failed on their way to the clinic. Those tests that are in clinical use — including prostate-specific antigen (PSA) for prostate cancer, mammogram-detected masses for breast cancer and CA-125 — fail to detect all cancers and sometimes ‘detect’ ones that aren’t there. Genomics, proteomics and other such technologies promised to help by finding combinations of markers that are more powerful and cancer-specific than individual ones, but that promise has not been realized. Researchers using such technologies have published studies on thousands of panels, suggesting that they can detect early-stage disease, guide patient treatment and monitor recurrence. But only a tiny number of such tests have reached the clinic — and none for the early detection of cancer, the biggest clinical challenge of all. “Much biomarker research has been done very badly for decades,” believes Lisa McShane, a biostatistician at the NCI in Rockville, Maryland. “Even when it was single markers. Now, as we’re moving up to multiple markers, all our bad habits are coming back to bite us in a big way.”

These habits have been thrown into the spotlight by the EDRN’s study, one of the largest and most systematic validation studies of biomarkers so far. It came just months after

a high-profile decision at Duke University in Durham, North Carolina, to suspend clinical trials of a genomics-based biomarker panel designed to direct chemotherapy in patients with breast cancer. A number of scientists had raised concerns about the Duke group’s data and analysis, and the trial was stopped after allegations came to light that the lead researcher, geneticist Anil Potti, had made false claims on his CV. Last September, the Institute of Medicine (IOM), part of the US National Academies, assembled a committee to discuss lessons for developing tests based on ‘omics’ technologies and bringing them to the clinic. “Why don’t we have assays out there, with this enormous promise?” Dan Hayes, a breast-cancer researcher at the University of Michigan in Ann Arbor asked researchers at the first IOM committee meeting in December 2010. “It’s either because these things just don’t work, or because we’ve used sloppy science to test them.”

It is too early to say whether either of these is true: the field is still young, and faces many challenges. It has drawn in many cancer biologists who are excited by the potential to translate their work to the clinic — but they sometimes lack the expertise or resources needed to pursue translational or clinical work. “A lot of novices came in. They get in without realizing that

the problem may be more complex than it appears,” says Eleftherios Diamandis, a clinical biochemist at the University of Toronto in Canada. And although most

experts agree that potential biomarkers for early cancer detection should be validated on samples taken before diagnosis — the stage at which the test would be used in the clinic — that is a step that few groups attempt and no biomarker for ovarian cancer has passed, as the EDRN study made clear. “Sometimes the glamour of the technology or the sheer volume of omics data seem to make investigators forget basic scientific principles,” said McShane at the IOM meeting. Mor agrees that the field has faced problems, and that it is important for markers to go through a careful process of design and validation, as he tried to do.

“There’s been an enormous amount of hype and promise,” sums up David Ransohoff, a cancer epidemiologist at the University of North Carolina in Chapel Hill. “But after 10 or 15 years of intense work in these fields, there’s simply not a lot to show for it. It’s important for the whole field to step back and look at what is wrong.”

#### MAKING A DIFFERENCE

Mor began his career in Israel, where he trained as a clinician at the Hebrew University of Jerusalem. But an experience in the final years of his oncology residency compelled him to change course. A young woman arrived at the hospital with ovarian cancer, a disease that kills some 140,000 women worldwide each year. The oncology team removed the woman’s ovaries and put her through several rounds of chemotherapy, which seemed to be successful. But 18 months later, she was back, her body riddled with tumours, and she soon died. “Chemotherapy didn’t do anything for her,”

**➔ NATURE.COM**  
A possible way  
forward for cancer  
biomarkers:  
[go.nature.com/icwtrue](http://go.nature.com/icwtrue)

IMAGES BY DU CANE MEDICAL IMAGING LTD/SPL

Mor recalls. “She was 29. She was a beautiful girl. An impressive girl. A medical student. And I never understood what happened to her.”

Mor decided to leave medicine, which had been unable to save her, for research, which one day might. He earned a PhD studying ovarian cancer at the Weizmann Institute of Science in Rehovot, Israel, before moving to Yale in 1997. He went on to start a programme called Discovery to Cure, aiming to speed cancer research to the clinic. The group began to build a bank of blood and tissue samples, including some from a Yale clinic for women with a high risk of ovarian cancer owing to a family history of the disease. “There was a lot of excitement around that time for finding proteins specific to cancer,” says Mor.

In 2003, David Ward, then a geneticist at Yale, contacted Mor. Ward had co-founded Molecular Staging, a company in New Haven that had developed a ‘high-throughput’ technique for quantifying multiple proteins in the blood using arrays of antibodies<sup>4</sup>. He asked whether he could use Mor’s samples to search for markers of early ovarian cancer.

Mor had never been involved with biomarker research — “I do biology of cancer, not biomarker development,” he says — but he signed up, intrigued by the clinical potential of the technology. Ward had scoured the literature for proteins that had been associated with ovarian-cancer growth and malignancy, and had come up with 169 candidates. Using the protein-quantification technique, Ward’s company screened blood samples in Mor’s tissue bank that came from two groups: women with newly diagnosed ovarian cancer who had been enrolled in Yale’s high-risk clinic, and women who had come to the hospital for routine gynaecological exams. Using additional cancer-patient samples, they whittled the list down to four proteins: leptin, prolactin, osteopontin and insulin-like growth factor II.

Mor worked to develop an algorithm that could automatically classify women as having cancer or not, depending on levels of these four proteins. When the team ran a new set of blood samples through the algorithm, they got astounding results. The test showed a sensitivity of 95% (meaning it correctly detected 95% of the ovarian-cancer cases) and a specificity of 95% (it erroneously classified only 5% of healthy people as having cancer). “I was delighted,” says Mor. On equivalent samples, CA-125 tests typically have a sensitivity of 70–80% and a specificity of around 95%. In May 2005, the findings were published in the *Proceedings of the National Academy of Sciences*

(PNAS), with Ward as a contributing author<sup>5</sup>.

Before publication, Mor helped the Yale Office of Cooperative Research to prepare a patent application. “A lot of companies expressed interest in licensing the panel,” says John Puziss, director of technology licensing at Yale. LabCorp licensed the test in 2006, as did Millipore, a biomanufacturing company based in Billerica, Massachusetts. (Mor says that the royalties he and his co-inventors received “were not a significant amount”.)

The test’s promising results had also caught the attention of researchers in the EDNRN, who were just putting together their validation study. Up to that point, most biomarkers for detecting early ovarian cancer had only been shown to distinguish patients with diagnosed cancer from healthy controls, but they are intended to detect the disease in women whose cancer is just budding, before symptoms develop. What the field needed was a ‘prospective’ study, run on blood samples from apparently healthy women, to see whether the biomarkers could pinpoint those who would later be diagnosed with ovarian cancer. Such samples, from large numbers of women who are tracked over months or years, are extremely difficult to come by.

#### PROBLEM DETECTION

The EDNRN found what was needed in the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial, sponsored and run by the NCI. Between 1992 and 2001, the trial had been collecting blood at regular intervals from 155,000 women and men, and screening them for cancer. By June 2006, 118 of the women had developed ovarian or closely related cancers, and the EDNRN researchers were now in a position to use them to evaluate the most promising biomarkers for early detection. Ziding Feng, a biostatistician at the Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, Washington, and coordinator of the EDNRN, visited Mor to discuss whether his panel of four proteins could be included in the study.

Mor was already in the process of refining the panel: he had more patient samples, and wanted to add more markers, including CA-125 and the protein macrophage migration inhibitory factor, to make the test more sensitive to cancer. LabCorp had been running his new samples on assay kits manufactured by Millipore. (Ward, meanwhile, had moved to the Nevada Cancer Institute in Las Vegas, and was not involved in data collection or analysis.)

When Mor showed Feng how he was analysing his recent data, Feng was troubled. Mor asked him to go through the new results himself, and Feng agreed to collaborate. “I do not do statistics,” says Mor. “That is not my field.” The researchers also added the six-protein panel to the EDNRN’s validation study.

Feng and Gary Longton, another statistician at the FHCRC, developed their own classification algorithms, and found that Mor’s test had a sensitivity of 95% and specificity of 99%. They also calculated the positive predictive value (PPV) of the test — the proportion of patients who the test would diagnose with the disease and do in fact have it. A high PPV means that few people will be misdiagnosed, which is crucial when screening healthy people.

Feng and Longton calculated the PPV at 6.5%, too low for the test to be of much use for screening. But separately, Mor was working with a different figure, of 99.3%. The huge disparity between the two values stemmed from the way that they calculated the figure and factored in the prevalence of ovarian cancer — an important variable in calculating the PPV. Following convention, Feng and Longton calculated the PPV using the accepted prevalence in postmenopausal women, 1 in 2,500 (0.04%). But Mor’s figure was calculated solely from the study population, in which the prevalence was 46%. “We calculated the PPV based on the population in the study, because we always intended the test for the high-risk population,” says Mor. “If you want to bring the test to the clinic, it has to be calculated based on the population you’re going to study,” he says, noting that other research studies work out the PPV for the study population in this way.

It’s a common mistake, believes McShane, who — like other statisticians — disagrees with Mor’s logic. “I see that a lot, but it is nowhere near the correct thing to do,” she says. Even in high-risk populations — women who are screened every year because of their family history or because they have tested positive for mutations in tumour-suppressor genes *BRCA1* or *BRCA2* — the prevalence is around 0.5%,

far below the 46% in Mor’s study population. Similar battles over the correct use of statistics litter the cancer-biomarker field, said researchers at the IOM meeting last year. “It’s the type of thing where non-statisticians think statisticians are being

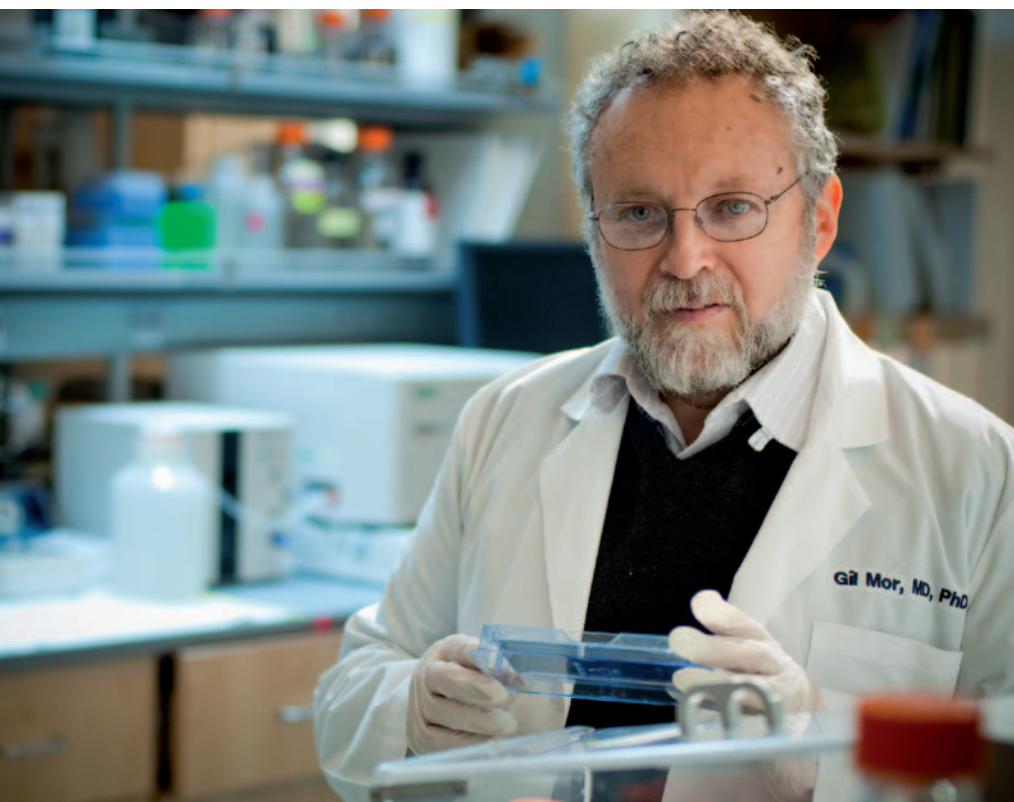
**“IT’S IMPORTANT FOR THE WHOLE FIELD TO STEP BACK AND LOOK AT WHAT IS WRONG.”**

uptight about something that’s not going to matter anyway,” says McShane.

Mor prepared a paper reporting the latest work. But when Feng and Longton saw the page proofs, they noticed that the PPV value was reported as 99.3%. They asked Mor to change it to the 6.5% that they had calculated, and to correct a few other typographical errors in the tables. “He agreed, so we signed off,” recalls Feng. But there was a miscommunication: Mor thought that Feng had agreed to the use of the high PPV, and that everyone approved of the final manuscript.

The paper was published online in *Clinical Cancer Research*<sup>6</sup> in February 2008, and to





Gil Mor is testing whether a panel of six proteins can detect ovarian cancer in women at high risk.

that “additional research is needed to validate the test’s effectiveness”. The paper in *Clinical Cancer Research* was also circulating at the Canary Foundation, a non-profit organization based in Palo Alto, California, that funds research on early cancer detection. Scientists there found other reasons for concern. One member, Nicole Urban, head of the Gynecologic Cancer Research Program at the FHCRC, had found that levels of prolactin, one of the proteins in the panel, are highly sensitive to stress — something very likely to affect women entering the clinic with symptoms of ovarian cancer<sup>7</sup>. After controlling for that, she says, “prolactin gave no signal at all for malignancy. It was useless.” Others pointed out that the high specificity and sensitivity figures reported in the paper’s conclusions, and trumpeted in Yale and OvaSure press releases, were not present in any of the tables or figures. And they bristled at the positive tone of the discussion, which stated that the test “will enhance the potential of treating ovarian cancer in its early stages and therefore, increases the successful treatment of the disease”.

“There were a lot of uncertainties, and evidence of biases,” says Martin McIntosh, who researches markers for early-stage ovarian cancer at the FHCRC, and is a member of the Canary group. “But the narrative only highlighted the best-performing analysis. It didn’t mention caveats.” Members of the Canary group wrote a letter to *Clinical Cancer Research*, describing some of their complaints. Meanwhile, Feng agreed to co-author a second letter, criticizing the paper even though he was a co-author.

The fuss was already reaching the FDA, which on 7 August 2008 sent a letter to LabCorp saying that the test “has not received adequate clinical validation, and may harm the public health”. A second letter, sent by the FDA on 29 September 2008, alleged that LabCorp did not have the necessary marketing clearance or approval for the test from the FDA. LabCorp replied to the FDA on 20 October, disagreeing with the agency’s assertions, but agreed to pull OvaSure from the market. It did so on 24 October 2008, just one day after *Clinical Cancer Research* published the critical letters from the Canary Foundation and Feng, as well as a third from the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia<sup>8–10</sup>. (Millipore continues to market the biomarker panel for use in research, not by patients.)

Mor was surprised by all three letters. In his published response<sup>11</sup>, he disputed some of the criticisms and wrote that any concerns about commercialization should be taken up with LabCorp. Stephen Anderson, vice-president of investor relations at LabCorp, says that OvaSure was not marketed as a test for detecting cancer recurrence, which was how some patients used it. He says that LabCorp “continues to believe OvaSure offers a

S. OGDEN

Feng’s shock it reported the high PPV. “You can imagine how upset I was when I saw it in the paper,” says Feng.

Feng called Mor. “I told him, those are errors, we told you those are not correct.” Feng also contacted the journal, the editor of which asked Mor to submit a correction to fix the PPV and the other typos. Mor agreed, adding the lower PPV as a footnote to the table and in a written correction.

A few weeks later, Feng received an e-mail with unwelcome news from a colleague: LabCorp was preparing to market the panel, and was “hopeful that this test will be available to women by the end of the year”.

“I was shocked,” says Feng. “I had no idea this was coming.” He thought that the markers should be validated further before they went to the clinic. In March 2008, Feng and Mor saw each other at a meeting in Washington DC. “I told him, face to face, you cannot do this,” says Feng. “You have to wait until after the PLCO validation. What you have done is early discovery. If validation does not support your earlier claim, you’re making a significant error.” Mor does not recall this encounter, but says that Feng’s “role was to analyse the data, not to make judgements of a company decision”.

Now, Mor says that if he were preparing the paper again, he would include both the low and high values for the PPV. And he vacillates about whether LabCorp’s decision to offer the test to women before it had undergone more validation studies was the right thing to do.

He says he thought that clinical use of the test might be a good way to do further validation. “It’s very difficult to do that on large numbers of patients,” he says. “It’s extremely expensive. The only way to do the study is if LabCorp started distributing the test and enrolling patients.” Mor notes that many tests, such as mammography, have been offered to patients as an aid to diagnosis even while data on the test are being collected. “Was it the right time? I don’t know,” he says.

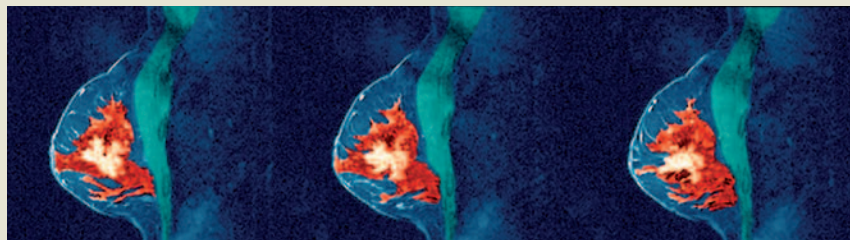
### CRITICAL BACKLASH

On 23 June 2008, LabCorp announced the availability of the OvaSure test, for between US\$220 and \$240. The press release said that it was being offered to women with a high risk of the disease, and quoted Mor as saying he was “pleased that this test is available to help physicians detect and treat ovarian cancer in its earliest stages”.

Excited chatter about the test spread through patient forums and support groups, but it was soon countered by cautionary tales. Jean McKibben, an ovarian-cancer survivor, rushed to take OvaSure on the first day it was available, and her results showed a 0.00 chance of cancer. A week later, scans showed that her cancer was back. She was crushed. “I wanted this to work so badly,” she wrote on a discussion board.

One week after LabCorp’s announcement, the Society of Gynecologic Oncologists in Chicago, Illinois, released a statement expressing concern about OvaSure, saying

## CASE STUDY

*The gene collection that could*

Genomic signatures measured in tissue samples can help to classify breast-cancer tumours.

When asked to name a successful cancer test that is based on multiple genes or proteins, many researchers point to Oncotype DX. The panel, which tests surgically removed breast tumours for the expression level of 21 genes to predict the likelihood of the cancer recurring, was developed by Genomic Health in Redwood City, California, and has been marketed since January 2004. It is used by roughly half the patients in the United States who have the most common type of breast cancer.

"A lot of biomarker research starts with getting some signature and then figuring out how to use it," says Richard Simon, a biostatistician at the US National Cancer Institute in Rockville, Maryland. But Genomic Health researchers took a different route, sitting down in 2000 with a group of oncologists and patient advocates and asking what question in cancer treatment they should address. Two years later, they nailed it down. At present, the majority

of women with the most common type of early breast cancer undergo chemotherapy after surgery, but only 15% are likely to have a recurrence. Was it possible to identify the crucial 15%, and spare the others from chemotherapy? The team set out to find a gene signature that could do the job. "The number one key to their success was starting with a well-defined, clinically relevant question," says Simon.

The researchers brought in Michael Walker, a statistician then based in Sunnyvale, California, to help design the studies from the outset. Walker says that it rarely works this way, and that often the statistician is only brought in after the data have been collected. By that time, biases and confounding factors may be hard-wired into the data. The team also put a high priority on using the right tissue samples in their initial studies. They decided early on that they wanted to use tumour tissue that had been fixed in formalin and embedded in paraffin

— the way it is prepared by the pathology lab after a tumour is removed — in the clinic and in all clinical trials.

Because of this, the team was able to validate the panel using samples that had already been collected in large clinical trials, rather than having to collect samples afresh.

"It's a poster child for one way to do clinical research," says David Ransohoff, a cancer epidemiologist at the University of North Carolina in Chapel Hill. But the test is hardly perfect. In January 2008, a group commissioned by the Centers for Disease Control and Prevention in Atlanta, Georgia, evaluated Oncotype DX. It found that the test results were reproducible and did well at predicting recurrence, but it was unclear whether the test was better than established risk factors, such as age, or standard molecular features of the tumour<sup>13</sup>. Results from a large, independent validation study, called TAILORx, are expected in 2015. **L.B.**

valuable tool for ovarian-cancer detection in conjunction with other diagnostic techniques", and that the assay is still in development. The company would not provide further comment.

## DOUBTS AND LESSONS

Since then, Mor has worked hard to validate his panel. He and Ward have completed a study on a much larger set of samples including many from women diagnosed in the

earliest stages of ovarian cancer<sup>12</sup>, and in which LabCorp again ran the assays. The test still performed well at distinguishing the patients from the healthy controls. Mor says he is puzzled by the PLCO trial results, and he hopes that further analysis of the trial data will help to explain why his biomarkers performed so poorly. He continues to express confidence in his panel, saying that the test could be most useful in high-risk populations, and when used regularly — every two to three

months — to monitor rising and falling levels of the biomarkers. But the whole experience has made him reluctant to pursue biomarker work much further. "I'm focusing on understanding cancer stem cells," he says.

Others say that's just as well. The panel's poor performance in the PLCO study makes critics question its usefulness in any group, even a high-risk one. McIntosh says that the PLCO study's damning conclusions should serve as a wake-up call. "The entire field has to cope with this," he says — including him, given that the most promising biomarkers discovered by his institution also failed to improve on CA-125 in the trial. "It's hugely disappointing."

The IOM committee, which is expected to release its results sometime in 2012, may help to find a way forward. At a meeting later this month, the members plan to draw lessons from the biomarker failures, as well as from the few success stories (see "The gene collection that could"). One of the most urgent lessons is the need to help researchers validate their biomarkers on appropriate samples before they reach the clinic. Feng says that the EDRN has been collecting its own high-quality tissue reference sets for ovarian, breast, lung, colon, liver and prostate cancers, from people who aren't yet showing symptoms and those in all stages of the disease. Investigators can apply to test their biomarkers on blinded tissue samples.

Until this type of testing becomes commonplace, there is no way of excluding the possibility that, as Hayes suggested at the IOM meeting, "these things just don't work" — particularly when it comes to picking up cancer early on.

"People keep talking about early-detection biomarkers as if they are a fact, and we only need to find them," says McIntosh, "when in reality their existence is a hypothesis that needs to be tested." ■ **SEE OUTLOOK P450**

**Lizzie Buchen** is a freelance writer in San Francisco, California.

1. Cramer, D. W. *et al. Cancer Prev. Res.* **4**, 365–374 (2011).
2. Zhu, C. S. *et al. Cancer Prev. Res.* **4**, 375–383 (2011).
3. Mai, P. L., Wentzensen, N. & Greene, M. H. *Cancer Prev. Res.* **4**, 303–306 (2011).
4. Schweitzer, B. *et al. Proc. Natl Acad. Sci. USA* **97**, 10113–10119 (2000).
5. Mor, G. *et al. Proc. Natl Acad. Sci. USA* **102**, 7677–7682 (2005).
6. Visintin, I. *et al. Clin. Cancer Res.* **14**, 1065–1072 (2008).
7. Thorpe, J. D. *et al. PLoS ONE* **2**, e1281 (2007).
8. Greene, M. H., Feng, Z. & Gail, M. H. *Clin. Cancer Res.* **14**, 7574 (2008).
9. McIntosh, M. *et al. Clin. Cancer Res.* **14**, 7574 (2008).
10. Coates, R. J., Kolor, K., Stewart, S. L. & Richardson, L. C. *Clin. Cancer Res.* **14**, 7575–7576 (2008).
11. Mor, G., Schwartz, P. E. & Yu, H. *Clin. Cancer Res.* **14**, 7577–7579 (2008).
12. Mor, G., Symanowski, J., Visintin, I., Birrer, M. & Ward, D. *Proc. Am. Assoc. Cancer Res. LB-224* (AACR, 2009).
13. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. *Gen. Med.* **11**, 66–73 (2009).





# THE HUNTING OF THE DARK

*The race to detect dark matter has yielded mostly confusion. But the larger, more sensitive detectors being built could change that picture soon.*

**F**or a substance that is utterly invisible, dark matter does a remarkably good job of making its presence felt. Astronomers have been compiling evidence for it since the 1930s, tracing how it shapes galaxies, galaxy clusters and even bigger cosmic structures by the inexorable force of its gravity.

Although its real nature is unknown, dark matter seems to outweigh the ordinary matter visible in stars and galaxies by roughly 5.5 to 1.

Down here on Earth, however, physicists struggling to answer the ‘what is it?’ question often feel like they’re chasing a ghost. Certainly, their detectors have been giving them a lot of strange and contradictory results. Two experiments are independently seeing what seems to be a flux of dark matter streaming through their apparatus. Another detector may have seen a handful of dark-matter particles last year — although the experimenters dismiss them as background noise. And yet another experiment has found no evidence for dark matter at all.

Fortunately, this confusion is likely to be temporary. Dark-matter detectors are roughly 1,000 times more sensitive to ultra-rare events than they were 20 years ago, and that should increase by another factor of 100 over the next decade, as physicists build bigger detectors and become more skilled at suppressing the background noise than can be confused with genuine signals (See ‘Dark-matter detectors’). “It would not be surprising if a year from now someone stood up and said we have done it, we’ve detected dark matter,” says Sean Carroll, a theoretical physicist at the California Institute of Technology in Pasadena. Other physicists give a more cautious estimate of five to ten years. Nonetheless, there is a palpable sense that the field is on the verge of something big.

Most of the attempts to detect dark matter directly have started from the assumption that the stuff is a haze of weakly interacting, massive particles (WIMPs) left over from the Big Bang. The ‘massive’ part would explain the gravity. And the ‘weakly interacting’ part would explain the invisibility: the WIMPs would flow through stars, planets and people in untold numbers, almost never hitting anything.

That assumption dictates the basic detection strategy: bring together a large target

BY ADAM MANN

mass of material; put it deep underground to shield it from cosmic rays and other radiation that could produce

misleading signals; then measure the recoil energy when a dark-matter particle finally hits an ordinary nucleus. The larger the mass of material, the more likely it is that a dark-matter particle will hit something.

Beyond those basics, setting up such an experiment requires a certain amount of guesswork. To have a significant recoil effect, for example, researchers need a target nucleus of roughly the same mass as the dark-matter particle they are seeking. It’s like watching for an invisible pool ball, says Jonathan Feng, a particle physicist at the University of California, Irvine. If the target nucleus is the equivalent of a bowling ball, the impact will barely move it. If, on the other hand, the target is the equivalent of a ping-pong ball, it will hardly be capable of deflecting the dark-matter particle, and so again there will be little energy transferred. What you want is another pool ball, Feng says.

## SUPERSYMMETRICAL WIMPS

Several dark-matter experiments have placed their bets on supersymmetry: a theory in which each particle in the standard model of physics would have a heavier, and so far unobserved, partner<sup>1</sup>. Supersymmetry predicts the existence of a WIMP called a neutralino, which would have exactly the right properties to account for the dark-matter distribution seen in the Universe. Its interactions would be feeble enough, yet its mass would be substantial — fifty to a few thousand times the mass of a proton.

One of the most highly regarded of the neutralino detection efforts is the XENON Dark Matter Search Experiment, located in the underground portion of the Gran Sasso National Laboratory near L’Aquila, Italy, and operated by a consortium of US and European universities. As its name suggests, the experiment’s detection medium is a tank of liquid xenon, which has a mass of just over 131 atomic mass units — close to ideal for detecting WIMPs at the lighter end of the supersymmetry range, which is by far the easiest place to start the search.

Photomultiplier tubes lining the inside of the XENON tank look for the characteristic flash of light called scintillation that would be generated if a xenon atom had recoiled from the impact of a WIMP. The XENON collaboration’s first detector, built in 2006, used around

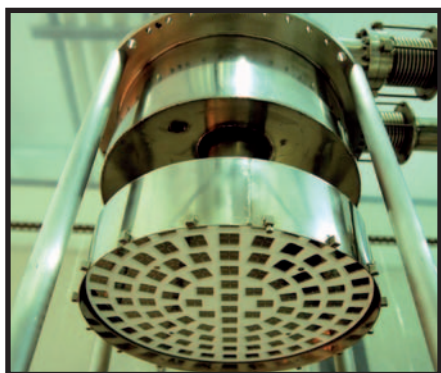
X-rays (pink) reveal ordinary matter in a galaxy cluster, and gravitational lensing of background galaxies (blue) maps dark matter.

X-RAY: NASA/CXC/CFA/M. MARKEVITCH ET AL.; OPTICAL: NASA/STSC/ MAGELLAN/UNIV. ARIZONA/ D. CLOWE ET AL.; LENSING MAP: NASA/STSC/ ESO WFI; MAGELLAN/UNIV. ARIZONA/ D. CLOWE ET AL.

## DARK-MATTER DETECTORS

Today's dark-matter detectors are 1,000 times more sensitive than their predecessors. The detectors coming on line will have bigger targets, making them even more sensitive.

XENON100 COLLABORATION; DAMA/LIBRA COLLABORATION



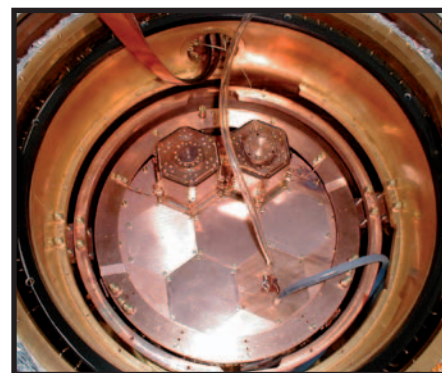
## XENON100

**LOCATION:** Gran Sasso, Italy  
**TARGET:** 161 kg liquid xenon  
**MASS RANGE:** 10–100 GeV  
**START DATE:** 2009  
**FINDINGS:** No detections



## DAMA/LIBRA

**LOCATION:** Gran Sasso, Italy  
**TARGET:** 250 kg sodium iodide crystal  
**MASS RANGE:** Light dark matter  
**START DATE:** 2008  
**FINDINGS:** Annual oscillations



## CDMSII

**LOCATION:** Soudan, Minnesota  
**TARGET:** 230 g germanium, 105 g silicon crystals  
**MASS RANGE:** 10–100 GeV  
**START DATE:** 2004  
**FINDINGS:** Two detections. Background noise?

CDMS COLLABORATION, FERMILAB

15 kilograms of xenon and found nothing that could not be attributed to background radiation. The team then upgraded to a bigger, more sensitive, 161-kilogram version in 2009, dubbed XENON100.

Although an initial 11-day data run on this detector still failed to find any particles<sup>2</sup>, that result was significant in itself: WIMPs with a mass of less than 100 gigaelectronvolts (GeV) should have shown up, says Laura Baudis, the physicist who leads the XENON group at the University of Zurich, Switzerland. Because they didn't, those lower masses could be ruled out. Unfortunately, the results from a subsequent, 100-day run remain obscure: the researchers are still struggling to deal with unexpectedly high levels of background radiation caused by trace contaminants in the xenon<sup>3</sup>.

## PURE AND SIMPLE

An experiment searching a similar mass range is the Cryogenic Dark Matter Search (CDMS) in the disused Soudan mine in northern Minnesota. As a detection medium, the CDMS team uses a collection of germanium and silicon crystals, which are among the only solid elements that can be made with high enough purity to be usable for detecting dark matter. When the detector is operating, these crystals — which are about 10 centimetres across — are cooled to a temperature of just 40 millikelvin, so any heat associated with a WIMP impact can be detected.

Now running its second-generation experiment, called CDMSII, the collaboration generated some excitement early last year when it reported two detections that could be interpreted as dark-matter signals<sup>4</sup>. Despite the hubbub, the team is reserved. "We don't claim this is significant; we see a lot of events at this low threshold and most of them are plausibly background," says Jeffrey Filippini, who works on the CDMS team from the California Institute of Technology. If those two events are discounted, the CDMS team gets much the same result as the XENON collaboration: null findings that effectively rule out low-mass WIMPs.

Yet the XENON and CDMS results contradict those from other experiments, the operators of which claim to have detected the very low-mass WIMPs ruled out by the first two. Perhaps the most intriguing, and most controversial, of these experiments is the Dark Matter Large Sodium Iodide Bulk for Rare Processes (DAMA/LIBRA), which shares space with XENON at Gran Sasso. DAMA works on the principle that the Sun's orbit around the centre of the Galaxy carries the Solar System through the invisible cosmic background of dark matter

at some 220 kilometres per second. So detectors on Earth should have dark matter flowing through them at that velocity, modulated by an annual variation of 30 kilometres per second as the planet orbits the Sun.

The DAMA team, which looks for the scintillation of recoil events inside sodium iodide crystals, claims to have followed just such a periodic dark-matter signal for thirteen years<sup>5</sup>. However, the crystals cannot distinguish between WIMPs and background events from ordinary radiation in the detector's surroundings, so this result depends on the assumption that background events occur at a constant rate that does not vary with the season. If that result is valid, it flies in the face of the XENON and CDMS findings.

"If the main signal was as big as they claim, we and other teams would have seen it," says Leo Stodolsky of the Max Planck Institute for Physics in Munich, Germany, who works on a collaboration called the Cryogenic Rare Event Search with Superconducting Thermometers (CRESST), also at Gran Sasso. Voicing a scepticism shared by many non-DAMA physicists, Stodolsky says that any number of seasonal processes could release subatomic particles that would mimic DAMA's results for dark matter, including something as simple as snow melting and refreezing in the mountains above the lab.

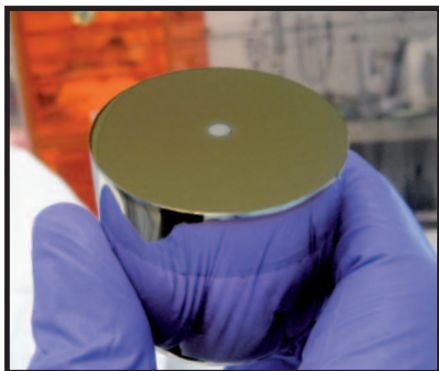
Further eroding DAMA's position is that no other dark-matter experiment is looking for a periodic signal, so its results cannot be directly replicated. Yet, despite the criticisms, the DAMA signal gets stronger each year. "DAMA has been very courageous," says Juan Collar, a physicist at the University of Chicago in Illinois. "They went out and made a claim" when most other physicists were still inclined to dismiss their results as background noise.

Collar leads an effort called Coherent Germanium Neutrino Technology (CoGeNT), whose detector sits near the CDMSII in the Soudan mine. CoGeNT uses germanium crystals tuned to detect incoming particles with much lower masses than those sought by its neighbour. It was originally intended to explore this range to rule out the existence of low-mass WIMPs, but its results only ended up making things murkier.

Around the time that the CDMSII reported its 'nearly nothing' findings, CoGeNT released data from its first 56 days of operations<sup>6</sup>. The results showed hundreds of particle events that could be interpreted as dark matter with a mass between 7 and 11 GeV.

These could also be the same particles that DAMA is detecting, but physicists have been quick to offer a more sober reading. "For CoGeNT,





### CoGeNT

**LOCATION:** Soudan, Minnesota  
**TARGET:** 500 g germanium crystal  
**MASS RANGE:** <10 GeV  
**START DATE:** 2004  
**FINDINGS:** Hundreds of detections



### LUX

**LOCATION:** Homestake, South Dakota  
**TARGET:** 350 kg liquid xenon  
**MASS RANGE:** 10–100 GeV  
**START DATE:** 2011  
**FINDINGS:** Not yet started



### XMASS

**LOCATION:** Kamioka, Japan  
**TARGET:** 1,000 kg liquid xenon  
**MASS RANGE:** 10–100 GeV  
**START DATE:** 2011  
**FINDINGS:** Not yet started

the signal and the background could easily be mistaken for the same thing,” says David Kaplan, a physicist at Johns Hopkins University in Baltimore, Maryland. The team has decided to wait a full year after its initial publication, to see if its findings show the same seasonal fluctuation as DAMA, before announcing any new results.

#### TOTAL ANNIHILATION

Meanwhile, a debate has broken out over another way to detect dark matter. One of the many oddities of dark-matter particles is that they can be their own antiparticles: put enough of them in one place, and they should start annihilating one another, producing  $\gamma$ -rays in the process. In particular, the centre of the Milky Way should be producing excess  $\gamma$ -radiation because dark matter is expected to concentrate there, says Dan Hooper, an astronomer at the Fermi National Accelerator Laboratory located near Batavia, Illinois. And Hooper claims to have found evidence for these  $\gamma$ -ray excesses in data from NASA's Fermi Gamma-ray Space Telescope<sup>7</sup>.

“If you were to ask what kind of a signal you would want to see with dark matter in the galactic centre, this would be what you expect,” says Neal Weiner, a theoretical physicist at New York University. The results are consistent with a dark-matter particle of 7.3–9.3 GeV, a range that fits well with the findings from both CoGeNT and DAMA.

Other researchers remain sceptical. “The galactic centre is so complicated that before you believe that you have dark-matter annihilation, you have to rule out all the options,” says Doug Finkbeiner, an astronomer at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. Finkbeiner points out that the signal could come from undetected pulsars — rapidly rotating neutron stars that produce copious amounts of high-energy radiation.

Still, Hooper's results have given researchers food for thought. “It's a case of too many coincidences,” says Collar. When findings from three detectors all start to point towards dark-matter particles of similar mass, he says, “you start to wonder if they're not coincidences any more”.

Such thinking has led theorists such as Feng to take a fresh look at all the results to see whether they can come up with a coherent idea of what dark matter might be. If CoGeNT and DAMA are right, says Feng, then

they are not detecting the expected dark-matter particle, the neutralino, as it should not be as light and interact as strongly as the results indicate. So perhaps dark matter is some very different particle — or perhaps the model of a single WIMP is not correct.

“If you look at the few per cent of the Universe that comprises us, it's quite complex,” says Philip Schuster, a physicist at the Perimeter Institute for Theoretical Physics in Waterloo, Canada, referring to the known ‘particle zoo’ predicted by the standard model that includes such oddities as muons, neutrinos and quarks. “It's a little insane to believe that the other 85% of the Universe would be so simple,” he says.

Along with his collaborators, Schuster is working to find evidence for a more complex theory of dark matter, called the ‘dark sector’. This sector could include multiple types of dark matter and a number of dark forces,

which, like ordinary matter, could combine to form dark atoms. It is being tested in an experiment called the A Prime Experiment (APEX), at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, which will accelerate a high-energy beam of electrons and search for relatively heavy force-carrying particles radiating from them. “It might tell us that the Uni-

verse is a lot broader than we suspect,” says Natalia Toro, a physicist who works at the Perimeter Institute and with Schuster on APEX.

The good news is that both the XENON100 and CoGeNT collaborations are expected to release their first full year's worth of data this year. And larger, more sensitive detectors, such as the Large Underground Xenon (LUX) and Xenon neutrino Mass (XMASS) detectors, are scheduled to start operations not long afterwards. “While we are in a ‘he said, she said’ situation now, it won't be like that indefinitely,” says Weiner. “We will have enough information to settle this in the next couple of years.” ■

*Adam Mann is a freelance writer in Washington DC.*

1. Waldrop, M. M. *Nature* **471**, 286–288 (2011).
2. XENON100 Collaboration. Preprint at <http://arxiv.org/abs/1005.0380> (2010).
3. Reich, E. S. *Nature* doi:10.1038/news.2011.125 (2011).
4. CDMSII Collaboration. *Science* **327**, 1619–1621 (2010).
5. Bernabei, R. *et al.* Preprint at <http://arxiv.org/abs/1002.1028> (2010).
6. Aalseth, C. E. *et al.* Preprint at <http://arxiv.org/abs/1002.4703> (2010).
7. Hooper, D. & Goonenough, L. Preprint at <http://arxiv.org/abs/1010.2752> (2010).

# COMMENT



**POLICY** Health experts need climate data to fight disease in Africa **p.440**

**EPIDEMIOLOGY** Tips for a happy, healthy life from long-running cohort study **p.443**

**EVOLUTION** From mammoths to mustangs, megafauna hold lessons for today **p.444**

**CORRESPONDENCE** Three responses to recent special on animal research **p.448**



F. GORO/TIME LIFE PICTURES/GETTY IMAGES

This 1961 effort to drill right through Earth's crust didn't succeed. Half a century on, geologists are ready to try again.

## Journey to the mantle of the Earth

On the 50th anniversary of the first attempt to drill into Earth's mantle, **Damon Teagle** and **Benoît Ildefonse** say that what was once science fiction is now possible.

Retrieving a sample of Earth's mantle has been an overarching ambition of the geoscience community for more than a century. In 1909, the Croatian meteorologist Andrija Mohorovičić noticed that seismic waves travelling below about 30 kilometres underground move faster than those above that depth, indicating a fundamental change in the composition and physical properties of the rocks. He had discovered the upper boundary of Earth's

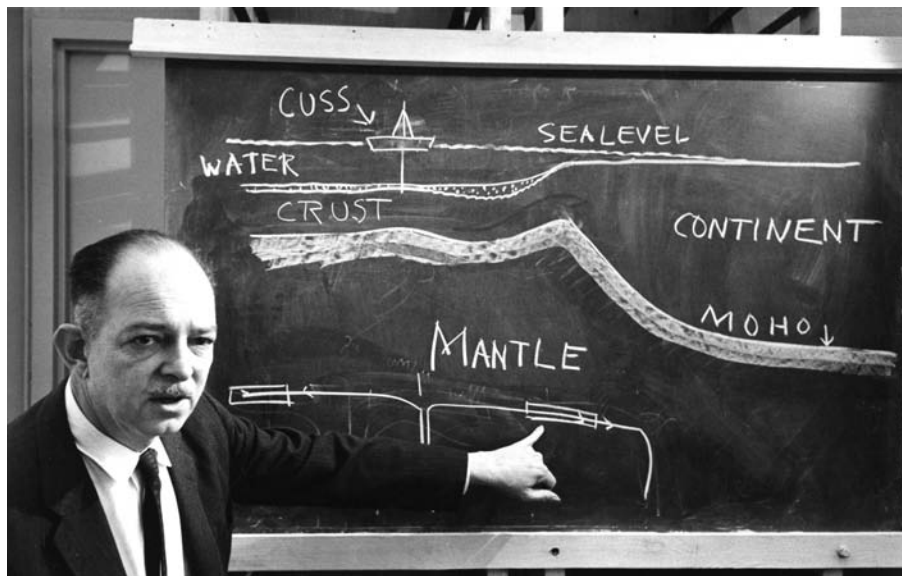
mantle, now known as the Mohorovičić discontinuity, or 'Moho' for short. This boundary marks the start of the bulk of Earth's interior, which extends from the base of Earth's crust — at 30–60 kilometres under the continents but just 6 kilometres under the thinner crust of the oceans — to the core 2,890 kilometres below.

Drilling down and retrieving samples directly from the mantle would provide scientists with a treasure trove comparable to

the Apollo lunar rocks, giving insight into the origins and evolution of our planet. But this has proved as difficult, perhaps more difficult, than going to the Moon. So far, no one has drilled deeper than about 2 kilometres into the ocean crust, or a third of the way to the mantle. The first effort to drill into the mantle, 'Project Mohole', foundered in a geopolitical quagmire and did not achieve its goal.

A new Mohole campaign is now under way, thanks to improved technology, ►





Harry Hess, a founding father of the theory of plate tectonics, explains Project Mohole.

► a better understanding of the rocks far below our feet and a deeper appreciation of the challenges of drilling through them. Over the next few years, geophysical surveys will be conducted at three Pacific Ocean locations that are in contention to be the site of the first deep hole to the mantle (see 'Drilling sites'). Drilling down to the mantle will require a huge amount of ship time and will be very expensive — far more expensive than a current single drilling expedition, although far cheaper than a Moonshot. But if funding can be found, and the scientific commitment maintained, drilling could begin within the decade, and be completed within 15 years. In the meantime, next month, we will be leading an expedition to the Pacific to bore further into the oceanic crust than ever before.

### INSPIRED IDEA

The first serious plans to drill down to the mantle were concocted in the late 1950s by a handful of post-war American geoscience grandees under the guise of the American Miscellaneous Society — an informal group of US National Academy of Science members, sometimes referred to as a 'drinking club'. The idea came primarily from Harry Hess, one of the founding fathers of the theory of plate tectonics, and Walter Munk, who pioneered studies of how winds drive ocean currents and explained why one side of the Moon is locked towards Earth. Frustrated by what they saw as a stream of worthy yet pedestrian research proposals in their field, they sought to undertake something more ambitious and innovative.

At a wine breakfast at Munk's home in La Jolla, California, on a Saturday morning in April 1957, they came up with Project Mohole, a scheme to drill, for the first time, right through Earth's crust and into the upper mantle.

Back then the nascent offshore petroleum industry had not yet begun to contemplate deep-water drilling. The Mohole project required the development of new technologies such as dynamic positioning, which would allow a drill ship to keep its position steady. The group obtained funding from the US National Science Foundation and commissioned the best ship available for the job: the drilling barge *CUSS 1*, named after the oil companies that had developed it, Continental, Union, Shell and Superior. Within four years of the project's proposal, propellers had been installed on the side of *CUSS 1* and a system developed that allowed these to keep the ship in position.

Between March and April 1961 scientists took their first core from the uppermost hard rock of the oceanic crust, or 'basement', off Guadalupe Island in the eastern Pacific Ocean, thanks to the daring and innovative engineering efforts of Willard Bascom and his colleagues. From beneath 3,800 metres of water and 170 metres of sediment, they pulled up a few metres of basalt, at a cost of US\$1.5 million (about \$40 million in 2009 dollars, in terms of its share of the total US economy). This remarkable accomplishment was reported in *Life* magazine (14 April 1961) by the novelist and amateur oceanographer John Steinbeck, who was aboard *CUSS 1* during these first operations.

This was the only ocean core that Project Mohole succeeded in drilling. After the expedition, the management of the project changed, some poor decisions were made about which drilling technologies to pursue, and costs spiralled out of control. In 1966, Project Mohole collapsed when the US Congress voted to cancel its funding.

Nevertheless, the project coincided with a growing acceptance of plate-tectonics theory. Interest in the formation and evolution of the

oceanic crust was booming. Project Mohole proved that scientific drilling into the ocean basement was possible. This contributed to the establishment, and continuance over four decades and running, of international collaboration in scientific ocean drilling. The Integrated Ocean Drilling Program (IODP; [www.iodp.org](http://www.iodp.org)) and its predecessors, the Deep Sea Drilling Project (DSDP) and the Ocean Drilling Program (ODP), form arguably the most successful, long-term international scientific collaboration in any field.

### THE DEEP FRONTIER

The mantle holds some 68% of the planet by mass. Its sheer volume makes an accurate knowledge of its composition and variability essential for understanding how Earth was formed and has evolved. Almost all of Earth's surface crust — the material that makes up the sea floor and the continents — originally came from the mantle.

Some pieces of the mantle have been thrust up to Earth's surface during tectonic mountain building, where they are available for study. Other mantle pieces, encased in lava, have been ejected from volcanoes, and sea-floor spreading has brought some to the ocean floor. These pieces show that the mantle is composed mainly of rocks called peridotites, made of magnesium-rich, silicon-poor minerals such as olivine and pyroxene. They also suggest, together with far-field seismic measurements, that the mantle's composition varies from place to place, but the extent of this variation remains unclear. The available samples have

**"Drilling to the mantle is the most challenging endeavour in the history of Earth science."**

all been chemically altered by the processes that brought them to the surface or by exposure to sea water. Concentrations of many of the key elements and isotopic tracers (including water, uranium, thorium, lithium, carbon, sulphur, silicon, potassium, the noble gases and the iron oxidation state) that might be useful in reconstructing Earth's evolution are highly labile. A few kilograms of fresh peridotite from beneath the crust would provide a wealth of new information.

Getting to the mantle requires drilling through a full section of oceanic crust, which will also be a boon for geologists. The formation of crust is the foundation of the plate-tectonic cycle. It is the main mechanism by which heat and material is dredged up from the interior of Earth, resurfacing some 60% of our planet every 200 million years. Currently, the thermal, chemical, and perhaps biological, exchanges occurring deep in the oceanic crust remain poorly understood because of the lack of direct observations *in situ*.

The technology to drill a hole a few inches wide through 6 kilometres of crust now exists or is feasible to develop<sup>1,2</sup>. A promising candidate technology is on the Japanese drilling ship *Chikyu*, launched in 2002. The vessel has a riser system: an outer pipe surrounds the drill string — the steel pipe through which cores are recovered. The drilling mud and cuttings are returned up to the vessel in the space between the two pipes. This helps to recycle the drilling mud, control its physical properties and the pressure within the drill hole and helps to stabilize the borehole walls. It also means that cuttings can be evaluated for scientific purposes. *Chikyu* is a giant ship, capable of carrying 10 kilometres of drilling pipes, and is equipped for riser drilling in 2.5 kilometres of water.

Over the next decade, researchers and engineers will have to design and develop new drill bits, lubricants and wireline instruments to make coring into the mantle possible, at pressures as high as 2 kilobars and temperatures up to 300 °C, and beneath about 4 kilometres of water. In particular, this will require a riser system capable of going deeper than *Chikyu*'s current equipment, or a different mud-circulation system, with part of the equipment installed on the sea floor.

To find the best site for the new Project Mohole, a large number of factors need to be considered. Ideally, it would be in the shallowest possible water overlying oceanic crust, which means going as close as possible to the mid-ocean ridge where new crust is formed. It should also be in the coldest possible crust, which is away from the ridge. These constraints limit the possible sites to three — off the coasts of Hawaii, Baja California and Costa Rica, respectively. All have pros and cons. The site near Hawaii, for example, is the coolest, but also the deepest,

and close to recent volcanic activity that might have chemically altered the mantle and perturbed the overlying crust. All the sites are in the Pacific Ocean because the crust there is formed faster than in other oceans, which makes for the simplest, most uniform basement architecture. Seismic and geological studies hint that fast-spreading ocean crust should be relatively uniform and conform most closely to textbook models. We hope to find material that conforms to these models of a simple crust: a layer cake of rock types called lavas, dikes and gabbros.

Meanwhile, the ocean-drilling community continues to core as deeply into the crust as is feasible using the conventional, non-riser technology available on the drilling vessel *JOIDES Resolution* — the recently rebuilt workhorse of the IODP.

### THE LOWER CRUST

Fifty years after John Steinbeck sailed on *CUSS 1* with the pioneers of ocean-crust drilling, we are the co-chief scientists on an expedition to obtain for the first time a section of the lower oceanic crust — the material lying just above the mantle. IODP Expedition 335 is due to sail from 13 April until 3 June<sup>3</sup>.

The site chosen for this mission is on the Cocos plate off the coast of Costa Rica (ODP site 1256). This site is in ocean crust that formed superfast — at more than 20 centimetres a year, much faster than any present-day crust formation. That makes the upper crust there much thinner than elsewhere, so it is possible to reach the lower portions without having to drill very deep. Three previous expeditions to Hole 1256D have drilled down to more than 1.5 kilometres below the sea floor, into the transition zone between dikes and gabbros<sup>4,5</sup>.

This spring we plan to deepen the hole

at least another 400 metres, and recover for the first time gabbros from the lower crust, which will be the deepest types of rock ever extracted from beneath the sea floor. (The hole itself is not the deepest ever drilled into the sea floor; that was Hole 504B, which reached 2,111 metres below the bottom of the eastern Pacific off Colombia.)

The mission should help to settle many debates: how crust is formed at mid-ocean ridges; how magma from the mantle is intruded into the lower crust; the geometry and vigour of how sea water can pull heat from the lower oceanic crust; and the contribution of the lower crust to marine magnetic anomalies. We will still be 3.5 kilometres shy of the Moho, but the project will provide further impetus for, and confidence in, deep ocean crust drilling, as well as crucial information for planning a full penetration through the Moho and into the upper mantle.

Drilling to the mantle is the most challenging endeavour in the history of Earth science. It will provide a legacy of fundamental scientific knowledge, and inspiration and training for the next generation of geoscientists, engineers and technologists. There is a surprising level of interest from the world's media and engagement by the general public in this frontier endeavour. And it may be only the beginning of a bigger project. As the crust and the mantle are both likely to be different from place to place, we will ultimately want a number of such holes. That may seem a distant dream, but ultra-deep drilling will only get more routine and cheaper with time and experience.

As Harry Hess told a US National Academy of Science meeting in April 1958, when defending the first Mohole Project against detractors: "Perhaps it is true that we won't find out as much about the Earth's interior from one hole as we hope. To those who raise that objection I say, if there is not a first hole, there cannot be a second or a tenth or a hundredth hole. We must make a beginning." ■

**Damon Teagle** is at the National Oceanography Centre, University of Southampton, Southampton SO14 3ZH, UK, and **Benoît Ildefonse** is at Géosciences Montpellier, CNRS, Université Montpellier 2, 34095 Montpellier cedex 5, France. e-mails: damon.teagle@soton.ac.uk; benoit.ildefonse@um2.fr

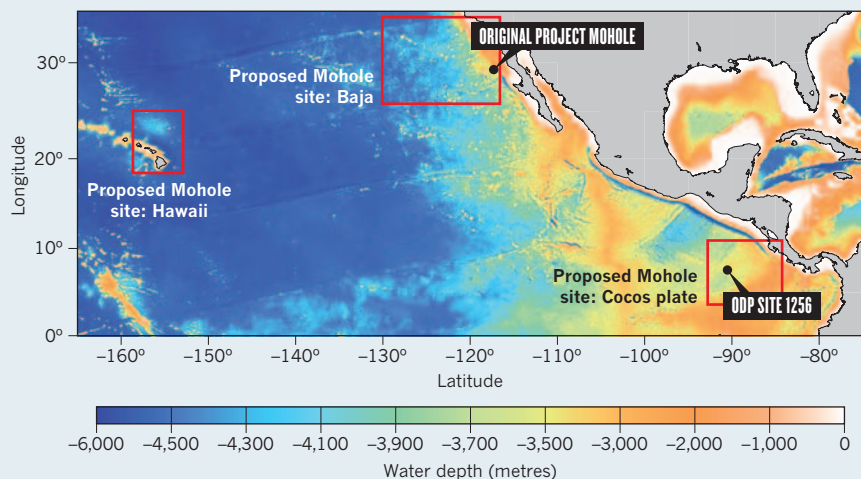
1. Ildefonse, B. et al. *Sci. Drill.* No. 10, 56–62 (2010).
2. Workshop Report Writing Group *Sci. Drill.* No. 11, 53–57 (in the press).
3. Teagle, D. A. H., Ildefonse, B. & Blum, P. *IODP Exp 335 Scientific Prospectus* doi:10.2204/iodp.sp.335.2010 (2010).
4. Wilson D. S. et al. *Science* **312**, 1016–1020 (2006).
5. Teagle D. A. H. et al. *Proc. Integrated Ocean Drilling Program* **309/312** Expedition Reports (IODP Management International, 2006).

Further reading accompanies this article online at [go.nature.com/b2ehus](http://go.nature.com/b2ehus).

SOURCE: REF. 1

### DRILLING SITES

Three areas are under consideration for drilling into the mantle. One includes the original Project Mohole drilling site. Another includes a site (ODP site 1256) where scientists will drill this year into the lower crust.







Ethiopian meteorologist Tufa Dinku working with Kenyan epidemiologist Judy Omumbo at Columbia University, New York.

# Africa needs climate data to fight disease

**Madeleine C. Thomson** and colleagues call on climate and health researchers, policy-makers and practitioners to work together to tackle infectious diseases.

Climate variability and change are a major concern for public health in Africa. The livelihoods of hundreds of millions of people there are dependent on rain-fed agriculture and seasonal water resources. Poor rural communities also suffer from undernutrition and bear the greatest burden of infectious diseases and natural disasters while having the least access to public-health services. Many of Africa's most important cities are on the coast and at risk of sea level rise. Without adequate infrastructure they are vulnerable to poor sanitation during floods and shortages of drinking water and loss of hydroelectric power during droughts. Rising temperatures, air pollutants and dust threaten to increase heat stress and respiratory disease.

The Fourth Assessment Report of the

Intergovernmental Panel on Climate Change predicts increased rainfall in eastern Africa over the coming century<sup>1</sup>. Yet there has been a region-wide drought over the past ten years<sup>2</sup>. Policy-makers want to know whether to prepare, short-term, for floods or droughts. They also need to know if the recent drying has aided malaria-control interventions in the region. But answering such questions is tricky.

Climate information is not readily available, so is rarely incorporated into development decisions. At the same time, few public-health institutions or practitioners are equipped to understand or manage the effects of a changing climate<sup>3</sup>, despite major advances in recent years in alerting the health community to its risks.

A dramatic improvement is needed in

the availability of relevant and reliable climate data and services, particularly in Africa, where vulnerability to climate is so high. Information — such as historical observations of temperature, ten-day satellite estimates of rainfall, the predicted start date of the rainy season or the likelihood of extreme temperatures in the coming season — should inform the management of all diseases sensitive to climate. These include: malaria, leishmaniasis, acute respiratory infections, intestinal helminths and diarrhoeal diseases. This information could also contribute to food security by providing, for example, early warning for agricultural and livestock pests and diseases.

The following must be put in place within the next decade: new partnerships between the public-health community and national

A. S. ORLING

meteorological agencies, space agencies and researchers; a governance structure that ensures data sharing between public and private agencies; a funding model that builds open-access climate databases; climate scientists focused on the delivery of quality products, tailored to user needs; health professionals trained to demand and use climate information; and evidence of the value of all this, relative to alternative investments in health.

### TRANSFORMATIONAL POWER

Good climate information, if freely available, could transform the way in which the health community does business. For example, it could improve health calendars for seasonal diseases. It could lead to better timing of the distribution of bed nets, local public awareness campaigns, and drugs with a short shelf life. Health professionals could be better prepared for the diseases that follow floods and storms, such as leptospirosis and cholera. It would also enable better mapping of regions and populations vulnerable to emerging health problems such as meningococcal meningitis epidemics, which favour the hot, dry and dusty Sahel, a region that may be expanding owing to climate and environmental change. On longer time scales, researchers could probe the drivers and potential recurrence of major climatic events, such as the devastating Ethiopian drought of 1984–85 (immortalized in the Western psyche by Bob Geldof's Band Aid).

On 4–6 April, in Addis Ababa, professionals — including policy-makers, practitioners, researchers, donors and the media — interested in using climate science to inform public-health decisions will meet at the Climate and Health in Africa: 10 Years On conference. Participants must reflect on the slow progress since the groundbreaking Climate Prediction and Disease/Health in Africa workshop in Bamako, Mali, in 1999. There, climate scientists committed to working directly with health practitioners and researchers. Participants must now agree a

road map for concrete action over the next decade.

The long-running debate on the likely effect of a warming climate on malaria in East Africa illustrates the problem. Over the past decade, numerous studies, with contradictory results, have attempted to explain the observed rise in malaria incidence in the western Kenyan highlands. For instance, peak monthly cases of malaria increased eightfold from the 1970s to the 2000s at the hospital that serves the Brooke Bond Farms (now Unilever Tea Kenya), tea estates near Kericho (with some tailing off in recent years)<sup>4</sup>.

Researchers who found no evidence of significant warming at this site during the same period concluded that only non-climatic factors (for example drug resistance, and local environmental change) could be driving the increased malaria incidence<sup>5</sup>. Those who found evidence of warming proposed that climate must play a part<sup>4</sup>. These conflicting results fuelled a heated and polarizing debate<sup>6</sup> in the malaria and climate-change literature (including this journal) and the media. The spat paralysed the discussion of climate and health at the highest policy level.

The limited access to daily observations of surface air temperature from meteorological stations, in quality-controlled, long, time series has constrained studies. They have relied heavily on short time series, used data of poor quality or ignored local ground observations in favour of spatially interpolated global data sets that could not provide meaningful results at local scale.

The value of the data held by national meteorological and hydrological services was made evident through a recent analysis of 30 years (1979–2009) of daily temperature and rainfall data from the Kericho meteorological station managed by the Kenya Meteorological Department; the data conform to World Meteorological Organization standards. The study<sup>7</sup> (the authors of which include S.J.C. and M.C.T.)

establishes that minimum and maximum temperatures, at this much-studied site, have been rising by about 0.2 °C a decade over the past 30 years. Furthermore, the study demonstrated the association between local temperatures in Kericho and sea surface and land surface temperature in the tropics — indicating the close relationship of the local Kericho climate to the larger climate system (see 'Going up').

Only now can researchers start to look properly for direct links between malaria incidence and climate variability and change in the area. We cannot yet say that warming helped the observed increase in malaria — although there is good reason to believe that it might<sup>4</sup>. But we can say that climate should not be dismissed as a potential driver in the area. The key concern is that, if there is a link, increased warming in the East African highlands may expose a largely naive population to malaria with devastating consequences.

That it took a decade to establish a robust analysis of climate trends in Kericho, a focus of so much controversy, points to a broader disconnect between those who need climate information and those who produce it. In the 1980s, African meteorological agencies were encouraged to sell their data to raise revenue to maintain their networks of meteorological stations. The agencies' services have understandably prioritized their primary client, the airline industry. Access for non-commercial purposes, including for malaria research, has been constrained by poor collaboration and high data fees, among other factors.

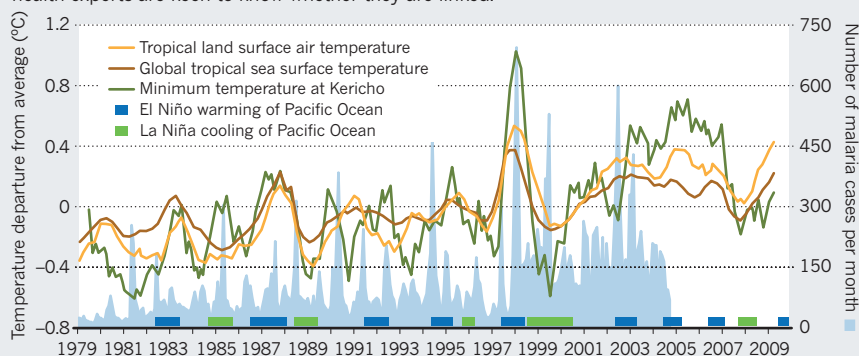
Instead, funding models are needed that recognize climate data as a resource for development — a classic public good that increases in value the more times the data are used.

### ETHIOPIAN PROMISE

The potential benefits of getting it right are considerable. Ethiopia, a country particularly vulnerable to the vagaries of climate, provides a promising example of what might be achieved. A new climate database there will draw on more than 600 national meteorological stations merged with 30 years of 10-day satellite data collected for rainfall monitoring. Generally fewer than 20 stations are available internationally via the Global Telecommunications System — the source of ground observations used in most available rainfall monitoring products. Ethiopia's National Meteorological Agency is getting technical support from the International Research Institute for Climate and Society (the IRI — where M.C.T., S.J.C. and S.E.Z. work) and the Tropical Applications of Meteorology using Satellite data (TAMSAT) research group at the University of Reading, UK. The effort is funded by Google.org, the philanthropic arm of Google, and lent

### GOING UP

Malaria incidence and temperatures have risen near Kericho in Kenya over the past 30 years; health experts are keen to know whether they are linked.



SOURCES: REF. 7; G. D. SHANKS ET AL. GO.NATURE.COM/CPN7KD



► technical support by the IRI through a cooperative agreement between Columbia University in New York (where the IRI is located) and the US National Oceanic and Atmospheric Administration.

This high-quality database will be used to create free-access climate reports tailored to the needs of the Ethiopian health community<sup>3</sup> and other development sectors, such as agriculture and water resources. It will also be used to improve the assessment of climate-sensitive interventions such as the indoor spraying programmes supported through the Roll Back Malaria initiative. One would expect such measures to work when conditions are least favourable to malaria transmission, for example, during a drought. The database will also help the development of local, seasonal climate forecasts — of unusually wet or dry conditions, say.

The Ethiopian climate database, the first of its kind, provides an opportunity to establish the value of climate information to improving health. Now that the system has been developed, the process can be more readily repeated in other countries. Doing so will build capacity where it is needed most — in the national meteorological agencies, regional climate centres and local universities<sup>9</sup>.

#### BETTER TOGETHER

Health professionals need skills to understand and interpret climate data, and to request new types of information or services. They also need to develop mechanisms to incorporate this information into current epidemiological approaches in a cost-effective manner.

One way forward is to target professional training and research in schools of public health. For instance, health-surveillance communities routinely monitor and prevent outbreaks and epidemics, through the analysis of current and historic epidemiological data. Where such events are climate-sensitive (for example Rift Valley fever epidemics) seasonal forecasts, meteorological information and satellite data could help map, monitor or anticipate changes in risk.

The Climate Information for Public Health Action Network, led by the IRI, and its associated training are steps in the right direction. The curriculum<sup>10</sup> enables climate and health experts to work together on common data sets and analyses, focusing the results on the needs of decision-makers. As a result, the African Field Epidemiology Network is exploring how climate information might be used in training for outbreak investigation.

The Climate for Development in Africa project was launched in October 2010. This is a joint initiative of the African Union Commission, the United Nations

Economic Commission for Africa and the African Development Bank. The project has a start-up fund of US\$136 million and a clear mandate from African heads of state to help fill key gaps in policy, practice, services and data across the continent. It is a daunting but necessary task. To achieve its development targets, the initiative will need to respond directly to the needs of climate-sensitive sectors, including health.

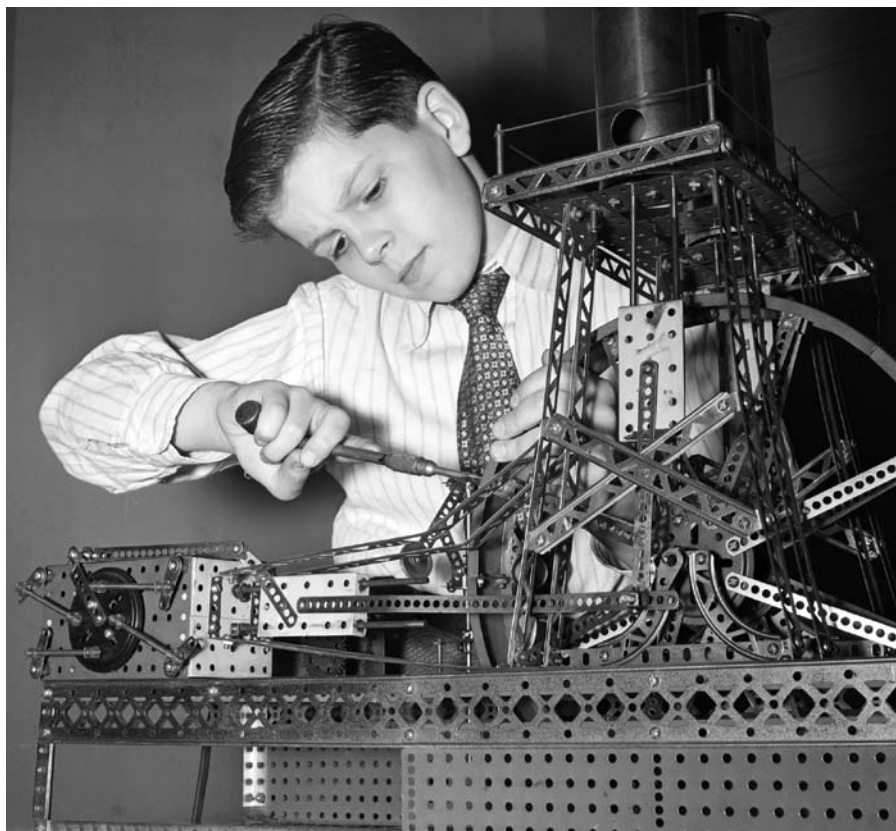
The global health community has worked for decades to get the resources necessary for effective control of diseases that affect poor people globally, especially malaria. Some people understandably fear that hard-won gains in political and financial support may be diluted, or worse derailed, by the climate-change agenda — especially in such aid-slashing times. But 'turf' anxieties are no reason for poor science. True interdisciplinarity requires more than fair-weather friends.

Climate is most important as a driver of infectious disease where and when control efforts are weak and societies are poor. Climate information can help to put resources where they are needed most. It is an essential additional layer of data for disease prevention, control and elimination.

History tells us that success against a single infectious disease such as malaria may be short lived if we are over-reliant on too few controls and lose a broad understanding of the disease. Rather than pursuing parallel agendas, the climate and health communities must work together now to deliver measurable health improvements in Africa in the next ten years and beyond. ■

**Madeleine C. Thomson, Stephen J. Connor and Stephen E. Zebiak** are at the *International Research Institute for Climate and Society, the Earth Institute, Columbia University, Palisades, New York 10964, USA*. **Michel Jancloes** is at the *Health and Climate Foundation, Washington DC 20005, USA*. **Abere Mihretie** is at the *Anti-Malaria Association, Addis Ababa, Ethiopia*.  
e-mail: mthomson@iri.columbia.edu;

1. Parry, M. L. et al. (eds) *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).
2. Park Williams, A. & Funk, C. *Clim. Dynam.* doi:10.1007/s00382-010-0984-y (2011).
3. Connor, S. J. et al. *Proc. Environ. Sci.* **1**, 27–36 (2010).
4. Alonso, D., Bouma, M. J. & Pascual, M. *Proc. R. Soc. B* doi:10.1098/rspb.2010.2020 (2010).
5. Hay, S. I. et al. *Trends Parasitol.* **18**, 530–534 (2002).
6. Reiter, P. *Rev. Sci. Tech. OIE* **27**, 383–398 (2008).
7. Omumbo, J. A., Lyon, B., Waweru, S. M., Connor, S. J. & Thomson, M. C. *Malaria J.* **10**, 12 (2011).
8. Ghebreyesus, T. A. et al. *WMO Bulletin* **57**, 256–261 (2008).
9. Dinku, T. *Clim. Dev.* **2**, 9–13 (2010).
10. Cibrelus, L. & Mantilla, G. *Climate Information for Public Health: A Curriculum for Best Practices Putting Principles to Work*. IRI Technical Report 163 (2010); available at [go.nature.com/3rg2eq](http://go.nature.com/3rg2eq)



Lewis Terman's study found that people who were conscientious as children were likely to live longer.

## HEALTH

# A long, diligent life

A 90-year cohort study hints that personality plays a unexpected part in lifespan, finds **Marten Lagergren**.

What factors predict a long, healthy and successful life? In 1921, psychologist Lewis Terman at Stanford University in California embarked on an ambitious project to find out. He selected around 1,500 gifted children from schools in the state and followed them from the age of 11 into adulthood, collecting a variety of data to see what might predict later success and accomplishment. Continued after his death in 1956 by other Stanford researchers and still ongoing, Terman's project has become the world's longest-running longitudinal study.

Psychologists Howard Friedman and Leslie Martin have brought the work up to date by painstakingly collecting information on those subjects who have died. Using death certificates, they have determined the length of life and cause of death, opening up a range of new analyses. *The Longevity Project* summarizes their findings on how life circumstances link to health outcomes, albeit for

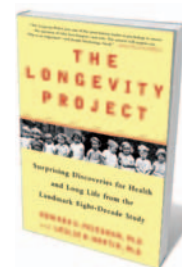
this select group. Some results are as expected, such as that smoking is bad for longevity. Others turn conventional wisdom on its head.

For example, working hard for long hours in a demanding job to achieve high status is better for your health and life expectancy than taking it easy and lacking ambition. Marriage is a blessing for men more than women; and men suffer more adverse health effects from divorce, perhaps turning to drink or drugs. The authors emphasize the benefits of an active social network — more common for women — as a buffer against life's harmful events. And they are critical of simple health advice, such as to jog or eat less fat, arguing that it is the whole approach to life that is essential, not the details. To give a person a list of health recommendations does not work, they point out, if the person cannot or does not follow them.

## ► NATURE.COM

For a review of two books on longevity science, see: [go.nature.com/kmdesk](http://go.nature.com/kmdesk)

Friedman and Martin explore how personality influences lifestyle choices using Terman's meticulous records of the character traits of the children he followed, as noted by parents and teachers. When the children were interviewed as adults 20 years later, the same qualities were evident. The best predictor of a long and healthy life turned out to be conscientiousness — the extent to which a child was prudent, dependable and persistent in the accomplishment of his or her goals.



## The Longevity Project: Surprising Discoveries for Health and Long Life from the Landmark Eight-Decade Study

HOWARD S. FRIEDMAN AND LESLIE R. MARTIN  
Hudson Street Press/  
Hay House: 2011.  
272 pp.  
\$25.95/£10.99

Conscientious people do more to protect their health and are less likely to engage in risky activities such as smoking, drinking or drug-taking, the study found. They also find their way to happier marriages, better friendships and optimum work situations. As a result, they are less likely to die from all causes.

Being physically active as a child is also a predictor of longevity, but only if that activity is maintained into and beyond middle age. The life-years gained by jogging may amount to no more than the time you spend doing it, the authors note. So we needn't all aim to run marathons; rather, we should just maintain an activity that we enjoy.

All this might seem to suggest that our fate is largely determined from childhood, but Friedman and Martin take a more constructive view. We can work with our personality to improve our health, they say, but it takes time for the benefits to accrue. You do not become conscientious overnight. It is the long-term, determined work of adopting and sticking to healthy habits and seeking good social environments and relationships that makes the difference. Later follow-up of Terman's subjects showed that conscientiousness in middle age and later counts almost as much as in childhood.

The authors provide self-tests for readers on a range of health-related factors such as catastrophizing (the tendency to always imagine the worst), life satisfaction, physical activity, marital happiness, job passion and accomplishment, and social-support network. They explain what these factors mean and provide guidance for improving one's lifestyle.

The difference in length of life between men and women has always intrigued epidemiologists and demographers. *The Longevity Project* adds a new twist by suggesting



that behaviour also influences the longer lifespan of women. In the study, children of either sex who were drawn to masculine careers (those shown by tests to be mostly preferred by men, such as being a mechanical engineer or pilot) had a shorter lifespan than those who preferred more feminine occupations (such as being an interior decorator or working with children). Thus, cultural dimensions may explain why life expectancy for the sexes differs over time and between countries and cultures.

There are caveats to this milestone study. One issue is that it was originally planned for a narrower purpose: to investigate predictions of career success and

**“Children of either sex who were drawn to masculine careers had a shorter lifespan than those who preferred more feminine occupations.”**

failure. Terman picked white pupils with high IQs from San Francisco schools, so the sample is not representative of the wider population. Conclusions cannot be drawn concerning minority

groups, educational level, social class or geographic area. The authors do their best to account for these limitations in their analyses.

Another problem is inevitable in any longitudinal study. Terman's subjects, who were born around 1910, had very different lives from ours. Many societal changes have occurred in the past century, particularly in gender roles. Terman's subjects, known as Termites, lived at a time when most women were expected to stay at home. The different life choices available today are likely to result in smaller gender differences in health and longevity.

*The Longevity Project* focuses mainly on the individual. The role of society in fostering good health and long life is seldom mentioned in the book, except when exposing the failure of current health propaganda. Despite ubiquitous recommendations to eat less and keep fit, obesity rates in the United States and in many other developed countries are soaring. The authors recognize that other studies are badly needed to examine the impacts of public policy on health and to develop more successful approaches. As they show in this excellent book, it will be a difficult task. But it is necessary. ■

**Marten Lagergren** is an assistant professor at the Stockholm Gerontology Research Center, Stockholm, Sweden.  
e-mail: [marten.lagergren@aldreacentrum.se](mailto:marten.lagergren@aldreacentrum.se)



MARK HALLETT/PALEOART/SPL

Giant ground sloths went extinct some 10,000 years ago, but could provide conservation lessons for today.

#### CONSERVATION

## After the auroch

Emma Marris is gripped by an account of our love-hate relationship with extinct megafauna.

It puzzles me that the many large, now extinct mammals of the Pleistocene Epoch have nowhere near the legions of fans claimed by dinosaurs. Mammals win the popularity contests among existing animals, yet few children can rattle off the weights and dietary habits of the gargantuan North American ground sloth *Megalonyx jeffersonii* or Australia's massive buck-toothed marsupial *Diprotodon optatum*. *Stegosaurus* gets all the love.

One fanciful explanation is that we have an abiding guilt for having killed them all off in our spear-hurling days. And it seems likely that human hunting played some part in many of these extinctions. In *Once and Future Giants*, biologist and journalist Sharon Levy lays out the evidence for this theory — and explores what this species drain can teach us now. The patterns and consequences of the Pleistocene die-offs can help us to predict how landscapes will change if we lose big mammals, and help us to spot warning signs of impending extinctions.

As we hesitantly take collective responsibility for these extinctions, we feel their loss more keenly. Today's 'wild' has diminished along with the megafauna. Spend enough time studying mastodons and moa, and even our most

rugged landscapes begin to look tame and denuded. North America's wolves and grizzlies no longer thrill; Yellowstone Park looks like a petting zoo. “We live in a highly abnormal world,” writes Levy, quoting US palaeoecologist David Burney. “We think of ground sloths and saber-toothed cats as peculiar and foreign, but it is the world of our own ancestry, the world our species evolved in.”

So, scientists and conservationists who can easily envision the landscapes of 13,000 years ago, just before the late Pleistocene extinctions, find themselves yearning for the past. They are starting to experiment with restoring these landscapes by introducing surrogates to fill long-vacant ecological roles — to graze, to browse, to kill, to knock over trees, even to terrify.

Levy recounts various rewilding experiments. Some have been intentional, such as the Pleistocene Park nature reserve in north-eastern Siberia, where rare native Yakutian



**Once and Future Giants: What Ice Age Extinctions Tell Us About the Fate of Earth's Largest Animals**  
SHARON LEVY  
Oxford University Press: 2011. 280 pp.  
\$24.95

➔ **NATURE.COM**  
For more on  
Pleistocene parks:  
[go.nature.com/kt4vnx](http://go.nature.com/kt4vnx)

horses roam. Others were accidental, such as the wild-mustang preserves of the American West. She reports on recent research supporting the notion that large animals are more than simply appealing — they can be major engineers of their ecosystems. Big predators such as the wolves of Yellowstone prevent herbivores from munching plant populations into oblivion and keep a lid on smaller predators. Big herbivores like the musk oxen of Greenland stop forests and weeds from overrunning the earth. They fertilize with their dung, and turn the earth with their big hooves.

Levy notes that many of the surrogates that conservationists use are the domesticated descendants of wild creatures. Specially bred cattle are used as proxies for extinct aurochs, the giant wild cattle that once roamed Europe, but Levy says that the modern cattle pale in comparison. Real aurochs — the kind painted by our ancestors in caves — were “longer of leg, bigger of brain, more graceful and fearless than their domesticated brethren”, she speculates.

The slightly mournful lesson of the book is this: any large animals we add to landscapes must be carefully managed. For example, condors reintroduced in the United States wear radio collars; wild mustangs are rounded up by the US government, dividing family groups and leaving excess animals held in pens. What differentiates such animals from pets?

To be truly wild, according to Levy, animals must have their numbers controlled by wild predators, not by humans. They must also live with fear. “The threat of a hungry carnivore lurking at the water hole is the essence of the truly wild horse,” she writes. And yet the idea of reintroducing predators — the key to wildness — is the most difficult to sell to local peoples around the world. Conservationists might love the thought of introducing African lions to the Great Plains in a bid to fill the gap left by the extinct American lion, but ranchers and rural residents understandably have qualms.

“We cannot raise the auroch, but its tamed descendent may yet fill a vital ecological niche,” concludes Levy in her examination of the increasing use of domestic cattle in conservation projects. Where once there were mammoths clashing tusks, giant short-faced kangaroos and woolly rhinoceroses, we now have Bessie the cow, grazing and fertilizing the soil and raising her head in vague interest as cars whizz past. It is one way of plugging the megafauna gap, but I long for the grandeur and strangeness of those lost giants. ■

**Emma Marris** is a writer based in Columbia, Missouri.  
e-mail: e.marris@gmail.com

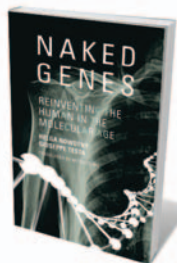
## Books in brief



### On Being: A Scientist's Exploration of the Great Questions of Existence

Peter Atkins OXFORD UNIVERSITY PRESS 152 pp. \$19.95 (2011)

Why are we here? Chemist and author Peter Atkins answers this big question succinctly and elegantly in this slim volume. Following in the footsteps of rationalists such as Richard Dawkins, he argues that we should find as much awe in the workings of science as we might in any god. Although he acknowledges the role of spiritual beliefs in society and the comfort they can bring to some, he finds greater solace in the scientific underpinnings of origins and endings, birth and death.



### Naked Genes: Reinventing the Human in the Molecular Age

Helga Nowotny and Giuseppe Testa MIT PRESS 192 pp. £18.95 (2011)

Advances in the life sciences have revealed many previously hidden aspects of biology, from the genes and proteins within cells to the developmental stages of the fetus. European Research Council president Helga Nowotny and stem-cell scientist Giuseppe Testa argue that these building blocks are not valueless, but are ‘naked’ blank canvasses that take on multiple meanings in different social contexts, from court rooms to parliaments. They assess how these varied perspectives influence attitudes to biotechnology in topics such as assisted reproduction and personalized medicine.



### Pox: An American History (Penguin History of American Life)

Michael Willrich PENGUIN PRESS 400 pp. \$27.95 (2011)

Attitudes to public-health interventions have not changed much in the past 100 years, explains historian Michael Willrich. He describes how measures at the turn of the last century to stem the spread of a smallpox epidemic in the United States — using quarantines, pesthouses and ‘virus squads’ — were met with suspicion and popular resistance despite their success. A well-organized anti-vaccination movement sprang up to champion personal choice over powerful government, resulting in the disputed political landscape around inoculation that is familiar today.



### Beyond the Finite: The Sublime in Art and Science

Edited by Roald Hoffmann and Iain Boyd Whyte OXFORD UNIVERSITY PRESS 208 pp. \$24.95 (2011)

How should we depict protein folding or negative mass? Scientists must create new imagery to describe such natural concepts every day, and in that sense they have a lot in common with artists who attempt to display the sublime. Nine scholars of science and art convey their perspectives in this volume. From the beauty of images taken by the Hubble Space Telescope to quantum romanticism, the contributors touch on natural aesthetics in physics, neuroscience, chemistry, painting and music.



### Bird Watch: A Survey of Planet Earth's Changing Ecosystems

Martin Walters UNIVERSITY OF CHICAGO PRESS 256 pp. \$45 (2011)

Bird populations worldwide are threatened by climate change and environmental destruction. This illustrated survey, produced in cooperation with the global conservation partnership BirdLife International, documents all 1,227 endangered bird species on the Red List of the International Union for Conservation of Nature. Region by region, the book describes the birds' habitats and the environmental pressures on them, as well as charting conservation efforts and top birding sites around the globe.



## MEDICINE

# Clinical precision

W. F. Bynum enjoys a history of three revolutionary moments in health care.

Harvard physiologist Lawrence Henderson once remarked that at some time between 1900 and 1912, a random patient with a random disease, choosing a physician at random, had for the first time in history a better than 50:50 chance of profiting from the encounter.

Henderson's spread bet is not quoted in *The Making of Modern Medicine*, but it concurs with historian Michael Bliss's take on how, when and where clinical medicine became modern. In this slim volume, based on his 2008 Joanne Goodman lecture series at the University of Western Ontario in Canada, Bliss offers three case studies that chart the leap in physicians' capacity to deal with disease at the start of the twentieth century. Health care was revolutionized by advances in disease prevention, surgery and drug treatments that allowed management of chronic afflictions. Bliss's measured evaluation of the strengths and weaknesses of modern medicine is persuasive and clear.

His first case study is a smallpox epidemic in Montreal, Quebec, in 1885, which sickened more than 20,000 people and killed about 5,000. It could have been prevented: vaccination was almost a century old, and the dynamics of smallpox spread were well understood. But anti-vaccination sentiment was rife. Bliss describes how social factors led to the outbreak being met with a mix of fatalism and ignorance. Most of the victims were children who were poor, French-speaking and Catholic. Their religious leaders failed them: inoculation was viewed as an affront

to divine providence and an intrusion of the state on a helpless citizenry. Two uneducated physicians led the resistance, telling parents that the removal of children to the quarantine hospital was a death sentence. As a result, infected children were kept in the community and became sources of further spread.

It took ten months for the epidemic to burn itself out. Many religious leaders maintained that the episode was an act of God sent to a sinful people. But, Bliss explains, most medical observers and liberal journalists at the time recognized it as a catastrophe that should never have happened, caused by antiquated values that were being supplanted by the new creed of science.

Spearheading that scientific approach was the Johns Hopkins School of Medicine in Baltimore, Maryland, the beginnings of which Bliss relates in his second case study. Its first professor of medicine was William Osler, who had left McGill University in Montreal in 1884 — a year before the epidemic — and moved to Johns Hopkins in 1889 after a spell at the University of Pennsylvania in Philadelphia. Osler excelled as a teacher and medical role model. He is still revered as such.



**The Making of Modern Medicine: Turning Points in the Treatment of Disease**

MICHAEL BLISS  
University of Toronto  
Press: 2010. 112 pp.  
Can\$21.95

The foundation of the Johns Hopkins School was a turning point in American medicine: it demanded that its students had a sound educational grounding, and provided both the facilities and the directive for its staff to be at the forefront of research as well as teaching. Its excellence was recognized in a gold-standard award from educational reformer Abraham Flexner, whose famous reports on the condition of medical education in North America in 1910 and in Europe in 1912 pulled no punches.

Osler was the institution's star, but Bliss concentrates on his surgical colleague, Harvey Cushing, as the most significant innovator. Cushing pioneered modern neurosurgery, creating the field almost single-handedly. He arrived at Johns Hopkins as a resident under the school's founding professor of surgery, William Halsted. Bliss explains that much of the therapeutic power of medicine in those days stemmed from surgery, a point that Osler also acknowledged. Surgeons could cure appendicitis, gall or kidney stones and other conditions that physicians could merely manage. Cushing could even operate successfully on the brain. Teachers and students at Johns Hopkins knew they were at the forefront of a medical revolution.

The decades before and after 1900 witnessed several medical innovations, including aspirin. But nothing captured the headlines like Bliss's third case study: the extraction of insulin in 1921 by Frederick Banting, Charles Best and James Collip in J. J. R. McLeod's lab at the University of Toronto. Louis Pasteur's rabies vaccine in the 1880s and Wilhelm Röntgen's discovery of X-rays in 1895 were comparable media events. But the dramatic pictures of emaciated children on the verge of death suddenly putting on weight after injections of insulin seemed almost miraculous.

Insulin did not cure type 1 diabetes, but it transformed an acute, fatal disease into a chronic one compatible with many years of life. The parallels between the treatment of type 1 diabetes in the 1920s and the treatment of HIV in our own time are striking.

We should be grateful that many diseases can now be managed. Yet Bliss reflects that the gratitude of people receiving medical treatment, as enjoyed by physicians half a century ago, has collapsed. He attributes this to a misplaced faith in medicine in a secular society. The fatalism of the Montreal poor has been replaced by an ardent expectation that medical science can solve the problems of the human condition. As a result, more of us may empathize with the poet Alexander Pope's sober analysis from more than two centuries ago: "This long disease, my life." ■

W. F. Bynum is emeritus professor of the history of medicine at University College London, UK.  
e-mail: w.bynum@ucl.ac.uk



The Johns Hopkins School of Medicine, founded in 1893, pioneered the scientific method in health care.



Benedict Cumberbatch plays Frankenstein's creation in a London adaptation directed by Danny Boyle.

## THEATRE

# A monstrous tale

A new staging of *Frankenstein* plays up the monster but draws no morals about science, finds **Philip Ball**.

Mary Shelley's *Frankenstein* has been adapted and reinterpreted innumerable times since it was first published, anonymously, in 1818. Victor Frankenstein has become the archetypal mad scientist, unleashing powers he cannot control. The 'Franken-' label is attached to every new technology that seems to intervene in life, from genetic modification of crops to Craig Venter's synthetic microbe.

The latest staging of the tale is Nick Dear's adaptation for the National Theatre in London, directed by Danny Boyle, famous for films including *Trainspotting* (1996) and *Slumdog Millionaire* (2008). The production is spectacular, and intelligent choices were made in the structure of the play (if not always in the dialogue). By placing the creature at the heart of the performance (played by Jonny Lee Miller, on the night I saw it), Dear has wisely emphasized the human message of the story rather than attempting any ill-advised jibe at the hubris of contemporary science.

The first adaptations for the stage began as early as the 1820s. These transformed Frankenstein's creature into the now-familiar mute, shambling brute, at that time based on the stock theatrical character of the Wild Man. This personification continued in the first film adaptation in 1910, simply called *Frankenstein*. As well as the iconic screen version by James Whale in 1931 with Boris Karloff's creature, there have been countless other renditions, sequels, parodies (Mel Brooks's 1974

*Young Frankenstein*, *The Rocky Horror Picture Show* in 1975) and postmodern interpretations (Brian Aldiss's *Frankenstein Unbound*; Jonathan Cape, 1973).

Some might lament how the original novel has been distorted and vulgarized in that process. British literary critic Chris Baldick has a wiser perspective. He argues in his book *In Frankenstein's Shadow* (Oxford University Press, 1987) that in interpreting a myth, we must consider all its versions: "That series of adaptations, allusions, accretions, analogues, parodies, and plain misreadings which follows upon Mary Shelley's novel is not just a supplementary component of the myth; it is myth." After all, there is no definitive version of Shelley's story. She made small but significant changes in the third edition in 1831, emphasizing the Faustian themes of presumption and retribution on which the early stage versions insisted.

Critics still dispute what Shelley's message was meant to be. Far from offering a simplistic critique of scientific hubris, the story might instead echo Shelley's troubled family life. Her mother, the feminist and political radical Mary Wollstonecraft, died from complications after Mary's birth, and her father, William Godwin, all

## Frankenstein

ADAPTED BY NICK DEAR. DIRECTED BY DANNY BOYLE  
National Theatre, London.  
Until 2 May 2011.

but disowned her after she eloped to Europe with the poet Percy Shelley in 1814. She lost her first child, named Clara, that year, subsequently describing a dream in which the girl was reanimated. There is ample reason to believe Percy Shelley's statement that the central moral of *Frankenstein* is: "Treat a person ill, and he will become wicked."

If so, Dear's adaptation has returned to the essence of the tale: it focuses on the plight of the creature, whose lone and awkward 'birth' begins the play. We see how this mumbling wretch, spurned as a hideous thing by Victor, is reviled by society until finding refuge with the blind peasant De Lacey. The kindly old man teaches the creature how to speak and read using Milton's *Paradise Lost*, the story of Satan's Promethean challenge to heaven.

Eventually, De Lacey's son and daughter-in-law return from the fields and drive out the creature in horror, whereupon he burns them in their cottage. These scenes are the moral core of Shelley's novel, and in placing them so early, Dear signals that this is very much the monster's show. In fact, perhaps too much. For while the creature is the most fully realized, most sympathetic and inventive incarnation I have seen, Victor Frankenstein is left with little to do but recoil from him and neglect all his other duties: martial, filial and moral. It is clear from the outset who is the real monster.

Throughout the run of this production, the two lead actors — Benedict Cumberbatch and Miller — alternate the roles of Victor and his creature. This Döppelgänger theme is not a new one. In the 1927 British stage adaptation by Peggy Webling that formed the basis of Whale's movie, the creature appeared dressed like the scientist, who foreshadows the later elision of creator and creature by saying: "I call him by my own name — he is Frankenstein."

The scientific elements of the tale are skated over. Mary Shelley provided just enough hints for the informed reader to make the connection with Italian physician Luigi Galvani's then-recent work on electrophysiology. Dear has Frankenstein mention galvanism and electrochemistry, but that is as far as it goes. There is no serious attempt to make the play a comment on the "Promethean ambitions" of modern science, as Pope John Paul II called them in 2002. This is a relief, because the trope of a solitary experimenter exceeding the bounds of God and nature is no longer the relevant vehicle for a critique.

Whether Dear adds anything new to the legend — as Whale and arguably even Brooks did — is debatable. But it is good to be reminded that the novel may be read not so much as a Gothic tale of monstrosity and presumption, but as a comment on the consequences of how we treat one another. ■

**Philip Ball** is author of *Unnatural: The Heretical Idea of Making People*.  
e-mail: p.ball@btinternet.com

➔ **NATURE.COM**  
For a review of Philip Ball's *Unnatural*:  
[go.nature.com/jgmb15](http://go.nature.com/jgmb15)



# CORRESPONDENCE

## Scientists should cut waste too

Your call for scientists to rally for continued federal funding (*Nature* 470, 305; 2011) places no responsibility on them to reduce the \$1.3-trillion US budget deficit.

As many scientists depend on taxpayers' money for research, they have an obligation to reduce waste and inefficiency and to work within their means. Funding agencies cannot and should not continue to do business as usual.

For example, the National Institutes of Health (NIH) imposes a salary cap of \$199,700 for scientists; most other federal agencies do not. The 'indirect costs' claimed by academic institutions range from 55% to 60% of the total grant budget. This implies that the taxpayer will pay \$199,700 for an NIH-funded radiologist but \$398,571 if the post were funded by another agency. Also, 55–60 cents of every research dollar will be spent on administrative and facilities costs, even though buildings and utilities have been paid for many times over.

Unlike companies, non-profit academic institutions deliver a paltry return on taxpayers' investments. In 2010, after spending nearly \$3.1 billion of taxpayers' money on intramural research, the NIH received \$91.6 million in royalties and was issued with 134 patents. By contrast, in 2009 IBM spent \$6.5 billion on research and development, generated \$15.1 billion in revenue and was issued with 4,914 patents. **Matthew Kumar** *Mayo Clinic, Rochester, Minnesota, USA.* [mkumar@mayo.edu](mailto:mkumar@mayo.edu)

## Anthropology: it can be interdisciplinary

Adam Kuper and Jonathan Marks's gloomy portrait of

integrative, big-question research in anthropology (*Nature* 470, 166–168; 2011) does not square with the large body of literature that covers areas such as behavioural ecology, cultural evolution, cognitive anthropology, gender studies, cross-cultural economics, moral psychology and environmental change. Publishing this work in high-impact general science and focused interdisciplinary journals ensures wide attention beyond the discipline.

The Evolutionary Anthropology Society was created to cut across traditional anthropological divides. It has some 350 members drawn from biological, cultural and archaeological specialities. Other interdisciplinary scholarly associations are The Human Behavior and Evolution Society, the European Human Behaviour and Evolution Association, and the Society for Anthropological Sciences. Each has hundreds of members active in the kind of research the authors claim is scarce or lacking. Productive interdisciplinary centres, such as the Centre for the Evolution of Cultural Diversity based at University College London, also catalyse innovative research that integrates biological, cultural and archaeological perspectives.

We feel that a genuinely interdisciplinary field of human diversity is emerging, synthesizing ideas and data from the social and behavioural sciences with theory and modelling techniques from evolutionary biology and game theory. Unlike Kuper and Marks, we see ample evidence that this work features in current debates about cognition, altruism, economic behaviour and environmental degradation (see, for example, M. Borgerhoff Mulder *et al.* *Science* 326, 682–688; 2009).

**Eric Alden Smith** *on behalf of 30 co-signatories\**, *University of Washington, Seattle, USA.* [easmith@u.washington.edu](mailto:easmith@u.washington.edu)

**Michael Gurven** *University of California, Santa Barbara, USA.*  
**Monique Borgerhoff Mulder** *University of California, Davis, USA.* \*A full list of signatories is available online at <http://dx.doi.org/10.1038/471448b>.

## Anthropology: follow field primatologists

Field primatology is one area of anthropology in which a classical cross-disciplinary approach is thriving (*Nature* 470, 166–168; 2011).

Field primatologists search the archaeological record of tool-using primates to gain insight into their cultures and traditions. Similarly, researchers of primate communication have set up a linguistic framework to investigate its intricacies in the context of the evolution of human language and music.

Like Jane Goodall and Birute Galdikas, whose studies on the great apes could read as ethnographies of a human group, field primatologists embrace long-term participant observation, a hallmark of social anthropology.

With the decline of natural forests, primate populations are nearly all intimately linked with their human neighbours. Field primatologists study their interactions, balancing the need for primate conservation with the cultural practices of the humans on whom the animals depend.

They advise on issues such as bushmeat hunting, the pet trade and the evolution of diseases that affect both human and non-human primates. They join cultural anthropologists and local people in examining data on past distributions and recent local extinctions of non-human primates and other animals.

In short, field primatology is successfully retaining and expanding the spirit of anthropology.

**K. Anne-Isola Nekaris**,  
**Vincent Nijman** *Oxford Brookes University, Oxford, UK.* [vnijman@brookes.ac.uk](mailto:vnijman@brookes.ac.uk)  
**Laurie R. Godfrey** *University of Massachusetts, Amherst, USA.*

## Intolerance: UK chief scientist responds

Andy Stirling and Brian Wynne (*Nature* 471, 305; 2011) call respectively for a democratic approach to scepticism and for recognition that scientific evidence often forms only part of complex decisions. I agree with them on both counts.

Of course it is true that advancement is attained through criticism, scepticism and debate. But my point was that there can sometimes be a thin line between healthy scepticism and a cynical approach that ignores or distorts inconvenient evidence.

Where significant consensus exists on an issue, this has not always been made obvious; also, tokenistic opposing views can be presented in a way that exaggerates their support.

Clearly, the role of scientific evidence in decision-making must be considered in the wider political and social context. However, I make no apology for demanding that the fundamental evidence and weight of consensus in such cases is set out in a proper and fair way.

**John Beddington** *Chief Scientific Adviser to HM Government, Government Office for Science, London, UK.* [mpst.beddington@bis.gsi.gov.uk](mailto:mpst.beddington@bis.gsi.gov.uk)

## Negative results are published

Jonathan Schooler argues in favour of an open-access database of negative results (*Nature* 470, 437; 2011). But publishing such results in scientific journals is

advantageous for authors, who can then list them among their papers.

Several journals specifically publish negative results. I'm aware of the *Journal of Negative Results in Biomedicine*, the *Journal of Negative Results — Ecology and Evolutionary Biology* and the psychology *Journal of Articles in Support of the Null Hypothesis*. There is a forum in the *Journal of Universal Computer Sciences* for negative results, and *PLoS ONE* also publishes them. Several other such journals have come and gone; all, I think, are open access.

Even so, negative findings are still a low priority for publication, so we need to find ways to make publishing them more attractive.

**Bob O'Hara** *Biodiversity and Climate Research Centre, Frankfurt, Germany.*  
bohara@senckenberg.de

## Animal research: a personal lesson

Had I been a participant in your survey on animal-rights activism (*Nature* 470, 452–453; 2011), I would have replied that animal extremism once had a negative effect on me — but in an unexpected way.

I worked for many years as a primate researcher studying animal models of abnormal development. Two years after the publication of Peter Singer's *Animal Liberation* (New York Review/Random House; 1975), my lab was attacked and its rhesus monkeys released. The monkeys were all recaptured and none was seriously injured. I felt intimidated, insulted and furious at what I saw as anti-science stupidity.

My anger was such that I did not give a thought to the possibility that the perpetrators might have been infected with deadly herpes B virus from the monkeys. I failed to alert the emergency departments in the area about this lethal possibility.

For years, my fury blocked the self-reflection that is expected of any scientist who harms vulnerable animals for presumed human benefit.

I dismissed even reasonable ethical questions directed at me and my work. Eventually, however, I took up a fellowship at the Kennedy Institute of Ethics at Georgetown University in Washington DC, and at the National Institutes of Health Clinical Center, where I studied bioethics on the moral standing of animals. My intellect and sense of compassionate responsibility broadened; research ethics became my life's focus.

Healthy debate about animal research and the ethical and scientific issues involved must be encouraged, even in the face of hostility. We must also remember that it is unreasonable and inaccurate to label everyone who opposes animal experiments as 'extremists'.

**John P. Gluck** *University of New Mexico, Albuquerque, New Mexico, USA.*  
e-mail: jgluck@unm.edu

## Animal research: the peaceful approach

In your articles on animal activism ([www.nature.com/animalresearch](http://www.nature.com/animalresearch)), there was no mention of the many individuals and organizations who work

peacefully and legally to educate the public and policy-makers about the ethical and scientific issues surrounding the use of animals in research.

At the American Anti-Vivisection Society, we seek to bring about meaningful, long-term change for animals in laboratories through the development and use of high-quality, non-animal-based teaching, testing and research.

Founded in 1883, the society brings a long-term perspective on opposing views and tactics. Biomedical research lobby groups in the United States have for decades opposed modest improvements to animal welfare laws and convinced researchers that there is too much red tape surrounding animal work. Yet the use of the most common lab animals — rats and mice — remains unregulated in the United States, and there is almost no accountability to the public, even regarding how many of these animals are used.

The same lobby groups attempt to sully the terms 'animal rights' and 'activists' by amplifying the illegal and offensive actions of individuals who do not represent any of us (see, for example, [go.nature.com/bxabrm](http://go.nature.com/bxabrm)). The reality is that 'peaceful' activists

often drive public policy on social issues. This has been true for animal issues for several decades and includes improvements to the US federal Animal Welfare Act.

**Crystal Miller-Spiegel** *American Anti-Vivisection Society, Jenkintown, Pennsylvania, USA.*  
cmillerspiegel@aavs.org

## Animal research: replacing the lab rat

Your coverage of animal research ([www.nature.com/animalresearch](http://www.nature.com/animalresearch)) focuses on well-worn themes from proponents, but does offer a way forward.

British biologist Peter Medawar predicted years ago that the use of animals in research would some day be completely replaced by more innovative methods (*The Hope of Progress*, Methuen; 1972). And Colin Blakemore, an ardent defender of animal research, has repeatedly stated that: "Everyone hopes that a time will come when no animal is used at all." To translate these congruous perspectives into action, we need to develop the kind of proactive strategies that you call for.

The results of your poll (*Nature* 470, 452–453; 2011) indicate that some scientists might be ready to take this idea forward. Others are clearly not immune to the ethical tensions in animal research. Sadly, most feel that the polarized debate on animal research makes it difficult to express more nuanced views, presumably because they do not want to be perceived as giving ammunition to the extremists.

Medawar's vision to replace animal experimentation is a goal that is worthy of serious effort, for the sake of scientific innovation, ethical responsiveness and animal protection. We should not be deterred by either the scientific challenges or the actions of a handful of extremists.

**Martin Stephens** *The Humane Society of the United States, Washington DC, USA.*  
mstephens@hsus.org





## Overcoming catalytic bias

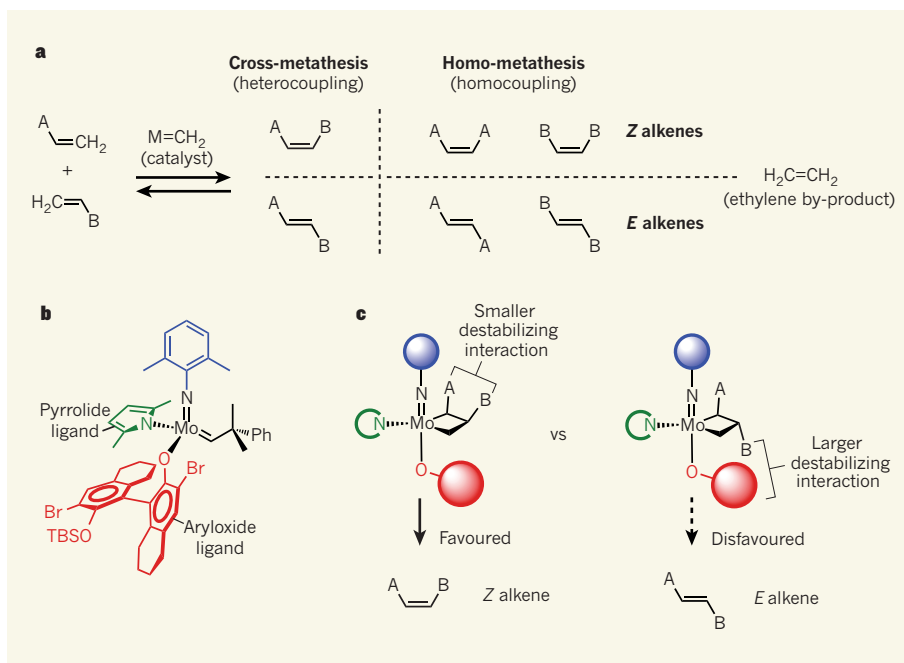
Metathesis reactions can be used to make carbon–carbon double bonds — bar one isomeric class. By using new catalysts and balancing out the stabilities of intermediates in the reaction, the elusive isomers can be made. [SEE ARTICLE P.461](#)

DAESUNG LEE

The alkene metathesis reaction, in which carbon–carbon double bonds (C=C bonds) are redistributed between alkene molecules, has had an enormous impact on chemical research and industry, as was recognized in 2005 with a Nobel prize<sup>1</sup>. But this remarkable reaction has an inherent limitation — it cannot generate the thermodynamically less stable isomers of the alkene products. On page 461 of this issue, Hoveyda and colleagues<sup>2</sup> report that they have overcome this problem using specially designed molybdenum catalysts. Their discovery opens up even more opportunities for the powerful metathesis reaction.

In alkene metathesis, two alkenes couple together to form a new alkene product along with an alkene by-product (typically ethylene, CH<sub>2</sub>=CH<sub>2</sub>; Fig. 1a). Conventional methods for making C=C bonds usually involve unstable reactants; by contrast, metathesis uses relatively stable alkenes as substrates, together with a metal alkylidene catalyst — a compound that contains a metal–carbon double bond. But metathesis tends to generate the more thermodynamically stable isomers (*E* isomers) of the products, rather than the less stable *Z* isomers. What's more, metathesis reactions are reversible, and *Z* alkenes are more prone to the reverse reaction than are *E* alkenes, reinforcing the tendency of metathesis to generate *E* alkenes. The preparation of *Z* alkenes by metathesis has therefore been a great challenge, except in a few cases<sup>3</sup>.

Hoveyda and co-workers<sup>2</sup> have met this challenge by using molybdenum alkylidene complexes as metathesis catalysts (Fig. 1b). These complexes overcome the inherent thermodynamic preference of the metathesis process, substituting it with a non-reversible reaction regime that selectively generates *Z* alkenes. The authors used a combination of electron-donating and electron-accepting ligands on the molybdenum atom of the complexes to improve catalytic performance, an effect that has been predicted by computational models<sup>4</sup>. Furthermore, they used a 'monodentate' ligand that makes the complexes structurally flexible, thereby allowing the catalysts to adapt to the structural strains imposed on them during



**Figure 1 | Selectivity in cross-metathesis.** **a**, Cross-metathesis (CM) reactions are intended to produce heterocoupled alkene products, but up to six possible products, including homocoupled ones, can actually form. *E* alkenes are generally produced, rather than the corresponding thermodynamically less stable *Z* alkenes. Ethylene is also produced as a by-product. A and B represent different chemical groups; M is the metal centre of the catalyst. **b**, Hoveyda and colleagues<sup>2</sup> report molybdenum (Mo) catalysts, one of which is shown here, that enable highly *Z*-selective CM reactions. Ph is a phenyl group; TBS is a tertiary butyldimethylsilyl group, (C<sub>4</sub>H<sub>9</sub>)Si(CH<sub>3</sub>)<sub>2</sub>. **c**, Two possible configurations are shown for an intermediate that forms during CM reactions using Hoveyda and co-workers' catalyst. Destabilizing interactions between A and B in the left-hand structure (which goes on to yield a *Z*-alkene product) are smaller than those between B and the aryloxide ligand in the right-hand structure (which yields the *E* alkene). The left-hand intermediate therefore forms preferentially, so that the CM reactions are *Z* selective.

the critical catalytic steps. These combined effects make the complexes especially effective catalysts for metathesis. The *Z* selectivity of the catalysts seems to be a consequence of the bulky, freely rotating monodentate ligand — in the reaction intermediate that leads to the formation of an alkene product, this ligand orients other bulky chemical groups so that they adopt the *Z* rather than the *E* arrangement in the product (Fig. 1c).

Although the new catalysts provided excellent reactivity and *Z* selectivity in metathesis reactions, the authors still needed to refine things further, especially for reactions involving two different alkenes. Such cross-metathesis (CM) reactions can generate up to six different products, because each reactant molecule

may react either with another molecule of itself (homocoupling) or with the other reactant (heterocoupling, the desired reaction), potentially yielding *E* or *Z* isomers in each case (Fig. 1a). For efficient heterocoupling, the reactivity of the reactants must be tuned appropriately. In a previous CM reaction<sup>5</sup> reported by the same research group, the authors tuned the reactivity of one of the reactants by using ring strain — an effect that destabilizes certain cyclic molecules. But no such obvious orchestration is possible in reactions of linear alkenes, such as those used in the present work<sup>2</sup>.

Hoveyda and colleagues realized that, to obtain heterocoupled *Z* alkenes from their CM reactions, and to prevent side reactions that regenerate starting materials, they would

have to obtain a delicate balance in the reactivities of the alkylidene intermediates formed during the reactions. To achieve this, they chose alkene reactants that included chemical groups that would bias the reactivities of the alkylidene intermediates through electronic and steric effects (repulsion that occurs when bulky groups are brought too closely together).

The first reactants they used were enol ethers — compounds in which a C=C bond is connected to the rest of the molecule by an oxygen atom (see Fig. 2 of the paper<sup>2</sup>). The C=C bonds in enol ethers have a higher electron density than those in simple alkenes, which makes them more reactive for metathesis. The authors found that, for the best conversion of reactants to products and for maximum yields of *Z* alkenes, the number of moles of enol ether used in the reactions should be ten times that of the other alkene reactant. By this means, they obtained high yields of heterocoupled products in reactions of enol ethers with simple alkenes, with exceptionally good *Z* selectivity.

Why are the reactions so successful? During metathesis reactions, a key intermediate known as a methylidene is produced. This methylidene can initiate reverse-CM reactions with *Z*-alkene products and generate unwanted *E* isomers. But in Hoveyda and colleagues' reactions, the methylidene reacts mainly with the enol ether reactants because there is an excess of them and because they are more reactive. What's more, the electronic properties of the resulting alkylidene intermediate are such that it preferentially reacts with an alkene rather than with another electron-rich enol ether. Heterocoupling is therefore the main reaction, rather than homocoupling between enol ethers.

The enol ether reactants used by the authors were cheap and readily available, but some enol ethers are much more expensive, in which case using a large excess of them would be undesirable. Fortunately, Hoveyda *et al.* found that unwanted methylidene formation can be minimized if the volatile ethylene by-product is removed by evaporation as it is produced. In this way, the authors maintained high *Z* selectivity in their reactions using just twice as much enol ether as alkene. They demonstrated the effectiveness of this protocol by using it to synthesize an unusual lipid containing a *Z* enol ether substructure.

The second reactants examined by the authors were allylic amides — compounds in which an amide group is attached at the carbon atom adjacent to a C=C bond (see Fig. 3 of the paper<sup>2</sup>). Unlike enol ethers, the C=C bonds in allylic amides are electronically similar to those in simple alkenes, making homocoupling between the amides a potential problem. But Hoveyda *et al.* used a compound that incorporates a bulky amide group, which prevents such homocoupling<sup>6</sup> — although homocoupling between the other

alkene molecules in the reaction could still be a problem. Nevertheless, the authors obtained excellent yields and *Z* selectivities in their CM reactions of allylic amides with alkenes.

The factor that determines the final product in these reactions is the relatively slow reverse CM of the heterocoupled product that is formed initially, compared with the faster reverse reaction of homocoupled alkenes. Removal of the ethylene by-product from the reactions was critical for efficient *Z*-selective CM, because this minimized the formation of methylidene species that convert *Z* isomers of products to their *E* isomers. The authors went on to use CM of an allylic amide with an alkene as a key step in the synthesis of the antitumour agent KRN7000. The use of this reaction allowed the agent to be made in just nine steps — the shortest synthetic sequence for the compound reported so far, underlining the effectiveness of the method.

Several aspects of these new metathesis reactions<sup>2</sup> remain to be further refined: the range of alkene substrates that can be used should be broadened beyond enol ethers and allylic amides, for example, and ways should be

found to avoid using an excess of one of the reagents. The conversion of starting materials to products in the reactions is currently not complete, so achieving complete conversion and higher yields of products without sacrificing the *Z* selectivity is also desirable. Nevertheless, these *Z*-selective CM reactions are highly promising and will potentially be of use for the preparation of numerous compounds, with far-reaching consequences for the future of metathesis chemistry. ■

Daesung Lee is in the Department of Chemistry, University of Illinois at Chicago, Chicago, Illinois 60607-7061, USA.  
e-mail: dsunglee@uic.edu

1. [http://nobelprize.org/nobel\\_prizes/chemistry/laureates/2005](http://nobelprize.org/nobel_prizes/chemistry/laureates/2005)
2. Meek, S. J., O'Brien, R. V., Llaveria, J., Schrock, R. R. & Hoveyda, A. H. *Nature* **471**, 461–466 (2011).
3. Crowe, W. E. & Goldberg, D. R. *J. Am. Chem. Soc.* **117**, 5162–5163 (1995).
4. Solans-Monfort, X., Clot, E., Copéret, C. & Eisenstein, O. *J. Am. Chem. Soc.* **127**, 14015–14025 (2005).
5. Ibrahim, I., Yu, M., Schrock, R. R. & Hoveyda, A. H. *J. Am. Chem. Soc.* **131**, 3844–3845 (2009).
6. Chatterjee, A. K., Choi, T.-L., Sanders, D. P. & Grubbs, R. H. *J. Am. Chem. Soc.* **125**, 11360–11370 (2003).

#### REPRODUCTIVE BIOLOGY

## In vitro sperm maturation

**Anticancer therapies can impair male fertility. Whereas men can opt to freeze their sperm before treatment, young boys don't produce mature sperm and so lack this choice. Work in mice offers hope for such patients. SEE LETTER P.504**

MARCO SEANDEL & SHAHIN RAFII

**T**he blueprint for producing mature, functional spermatozoa in a laboratory dish — all the way from stem cells to flagellated sperm — has eluded reproductive biologists for decades. With a view to the eventual *in vitro* production of human sperm for clinical use, the main criterion of normal gamete function is the ability to support fertilization, with the subsequent development of normal offspring. Reporting on page 504 of this issue, Sato *et al.*<sup>1</sup> meet this challenge in mice.

The process of spermatogenesis in mammals persists throughout almost all of adulthood. It starts with spermatogonial stem cells, which differentiate from type A spermatogonia into type B spermatogonia, and then into spermatocytes. The spermatocytes undergo meiotic cell division to form spermatids, and finally spermatozoa<sup>2</sup>. Complete maturation takes more than a month in most mammals.

Previous efforts to nurture sperm *in vitro*

have either managed to recapitulate only parts of this complex differentiation process or failed to demonstrate the production of normal, fertile offspring. Although the organ-culture methods developed in the 1960s allowed germ cells to progress to meiosis<sup>3</sup>, until now no methods could support the entire process.

Sato *et al.*<sup>1</sup> reasoned that the orchestration of cell-maturation signals during spermatogenesis should be achievable in a dish by providing nearly all of the cellular components found in the testes. To this end, they cultured testicular fragments to maintain the proper microenvironment for cell differentiation. This involved suspending fragments of immature testis on a semi-solid support, such as agarose, and partially bathing them in liquid. This organ-fragment culture system — referred to as the gas-liquid interface method — balances the delivery of nutrients from the culture medium to the maturing cells with the need for efficient gas exchange to maintain spermatogenesis for more than two months.

To visualize and score maturing germ cells





## 50 Years Ago

“The great majority of school children are not only robust and healthy, but are taller and heavier than their predecessors”, states Sir John Charles, the chief medical officer of the Ministry of Education, in his report for 1958 and 1959 ... Only about five per cent of school children contracted a notifiable infectious disease in the years under review. Tuberculosis continued to decline. Poliomyelitis reached its lowest level for thirteen years, and vaccination against it was undertaken vigorously everywhere ... The main cause of death among children is through accident ... In 1958, 869 children of 5–15 years died from accidents, including 395 from accidents involving motor vehicles, 174 from drowning, and 58 from burns and scalds.

From *Nature* 25 March 1961

## 100 Years Ago

I have just been told a very interesting story by Mr. James Day of this town. Many years ago he and his father ... noticed a fox come searching along the hedgerows ... they saw that he was collecting the sheep's wool caught on the thorns and brambles. When he had gathered a large bunch he went down to a pool ... and backed slowly brush first into the water, until he was all submerged except his nose and the bunch of wool, which he held in his mouth. He remained thus for a short time, and then let go of the wool, which floated away; then he came out, shook himself, and ran off. Much astonished at this strange proceeding, they took a shepherd's crook ... and pulled the wool out. They found that it was full of fleas, which, to save themselves from drowning, had crept up and up the fox's brush and body and head and into the wool, and that thus the wily fox had got rid of them. *T. McKenny Hughes*

From *Nature* 23 March 1911

under different conditions, the authors used a marker for spermatogenesis<sup>4</sup>: their mice were genetically engineered to express green fluorescent protein (GFP) under the control of regulatory elements for genes that are activated only when germ cells progress into meiosis and beyond. As expected, the tissues that they collected from the testes of newborn mice showed little or no baseline expression of GFP.

The authors had previously optimized<sup>4</sup> various parameters involved in *in vitro* sperm formation, including temperature and the choice of basic ingredients for the medium. Notably, fetal bovine serum (FBS) seemed to be an important component, because its absence resulted in negligible maturation — as evidenced by the lack of the GFP signal.

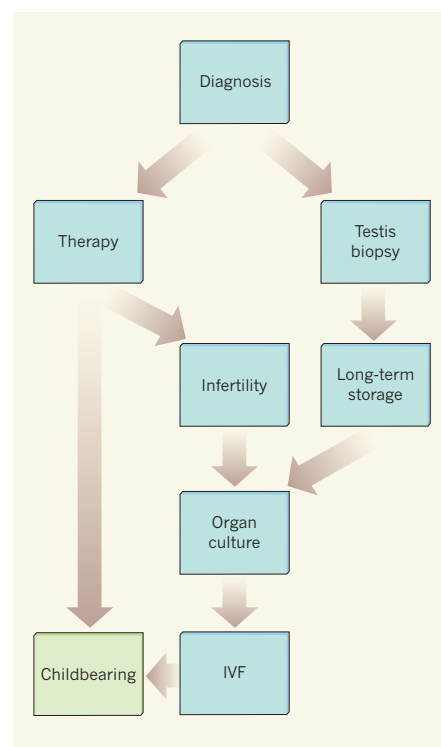
In their present work, Sato *et al.*<sup>1</sup> confirm the importance of FBS but, borrowing from the field of embryonic-stem-cell biology<sup>5</sup>, they find that an alternative to FBS known as knockout serum replacement (KSR) is even more effective. This finding is counter-intuitive, because KSR is commonly used to maintain stem cells in an undifferentiated state. A clue to the mechanism involved comes from the fact that the lipid-rich albumin component of KSR is itself highly effective in boosting differentiation.

The authors used *in vitro* fertilization (IVF) techniques to demonstrate the authenticity of the sperm collected from their cultures: they obtained male and female offspring that were themselves fertile.

The preservation of fertility is a major concern for patients requiring therapy, such as chemotherapy or radiation therapy, that can inadvertently destroy germ cells. In men, this problem can be mitigated by banking sperm before treatment. The solution is less straightforward in pre-pubescent boys. On the basis of pioneering work in animals by Brinster<sup>6</sup> and others, the idea of transplanting cryopreserved spermatogonial stem cells is a reasonable strategy, although it has not yet been rigorously assessed in humans. Furthermore, the technology for the long-term culture and expansion of human spermatogonial stem cells has not been standardized, nor has the safety of the approach been tested.

Sato and colleagues' results suggest a viable alternative. In this scheme, boys would undergo testicular biopsy before chemotherapy or radiation therapy, to obtain tissue for cryopreservation (Fig. 1). If infertility occurs, the testicular fragments could be thawed and sperm obtained from organ culture for IVF. Such a protocol would bypass the need for surgical spermatogonial stem-cell transplantation.

It remains unclear whether the success of this system is due to signalling molecules released by the germ cells themselves, or to molecules released by the surrounding somatic (non-germ) cells, or to both. Nonetheless, the integrity of somatic cells, especially Sertoli cells, seems to be essential. However, this is not



**Figure 1 | Possible scenario for preserving fertility, based on the new work<sup>1</sup>.** After a boy has been diagnosed with cancer, a biopsy would be performed to obtain testicular fragments for long-term cryopreservation. Later in life, when the individual wants to start a family, fertility would be assessed, and if he cannot produce functional sperm, the stored tissues could be thawed for organ culture. Sperm formed by this *ex vivo* method could then be used for *in vitro* fertilization (IVF).

surprising, given that germ-cell maturation is known to depend on somatic-cell signals. In fact, even when embryonic stem cells have been used to produce germ cells *in vitro*<sup>7</sup>, signals contributed by differentiating testicular somatic cells in the culture seem to be required.

The exact nature of the external signals that enhance sperm maturation is not the only remaining mystery. It is also not known whether the resulting offspring — especially those produced from cryopreserved tissue — are generally healthy. Indeed, fertility of the offspring is just a crude indicator of whether gametes are ‘normal’. Investigations should be made into whether the progeny Sato and co-workers generated by IVF are healthy in other ways (with respect to ageing, immune function, behaviour and so on).

As for the consequences of *in vitro* spermatogenesis at the molecular and cellular levels, previous data have indicated<sup>8</sup> that adverse epigenetic effects occur when cells, especially gametes, are maintained in culture. Whether DNA repair, which is essential for spermatogenesis *in vivo*, functions normally *in vitro* is another concern. Subtle genetic or epigenetic changes could be pivotal for the well-being of subsequent generations. These caveats aside,

the organ-culture approach<sup>1</sup> represents a crucial experimental advance along the thorny path to the clinical use of sperm developed *in vitro*. ■

**Marco Seandel** is in the Department of Surgery, and **Shahin Rafii** is in the Department of Genetic Medicine and at the Ansary Stem Cell Institute, Weill Cornell Medical College, New York, New York 10065, USA.  
e-mails: mseandel@med.cornell.edu;  
srafii@med.cornell.edu

## SCHIZOPHRENIA

# Zooming in on a gene

**Genome-wide association studies are often criticized for providing little insight of immediate physiological relevance. The finding of one such study, which implicates a signalling molecule in schizophrenia, is welcome news. SEE LETTER P.499**

HUGH D. PIGGINS

Schizophrenia affects between 3 and 7 in every 1,000 individuals<sup>1</sup>. Yet, as for most mental disorders, the origins and mechanisms of this multi-symptom psychiatric illness remain elusive. In this respect, Vacic and colleagues' report<sup>2</sup> on page 499 of this issue — the first genome-wide association study to implicate a single gene in this complex disease — is enlightening.

Throughout the late twentieth century, schizophrenia research focused mainly on alterations in the levels of the neurotransmitter dopamine and in particular of its receptors<sup>3</sup>. However, this proved to be only one piece of the puzzle.

Hopes were again raised with the emergence of tools that can detect and map genetic variations and alterations. One such modification, known as copy number variation (CNV), involves structural changes in the genome that result from disruptions, duplications or deletions of sections of DNA. Although researchers tend to look for common CNVs that are associated with complex disorders, large-scale and sensitive genome-wide association studies, in which thousands of genes in patient samples can be assessed, also allow the detection and mapping of rare CNVs and the genomic regions (loci) enriched in them.

One idea that is rapidly gaining momentum is that the accumulation of rare CNVs is associated with a high risk of developing schizophrenia and other mental disorders such as autism<sup>4</sup>. Indeed, CNVs on several chromosomal loci (microdeletions at loci 1q21.1, 3q29, 15q13.3 and 22q11.2, as well as microduplication at the 16p11.2 locus) are associated with the risk of schizophrenia<sup>4,5</sup>. Together, however, these CNVs account for only 2–4%

1. Sato, T. *et al. Nature* **471**, 504–507 (2011).
2. Russell, L. D., Ettlin, R. A., Hikim, A. P. S. & Clegg, E. D. *Histological and Histopathological Evaluation of the Testis* (Cache River Press, 1990).
3. Staub, C. J. *Androl.* **22**, 911–926 (2001).
4. Gohbara, A. *et al. Biol. Reprod.* **83**, 261–267 (2010).
5. Price, P. J., Goldsborough, M. D. & Tilkins, M. L. International patent application: PCT/US98/00467 (1998).
6. Brinster, R. L. & Zimmermann, J. W. *Proc. Natl Acad. Sci. USA* **91**, 11298–11302 (1994).
7. Kee, K., Angeles, V. T., Flores, M., Nguyen, H. N. & Reijo Pera, R. A. *Nature* **462**, 222–225 (2009).
8. Nayernia, K. *et al. Dev. Cell* **11**, 125–132 (2006).

of schizophrenia cases and do not extensively contribute to our understanding of the mechanism of this disorder.

Vacic *et al.*<sup>1</sup> carried out a genome-wide association study of 8,290 patients with schizophrenia. They find that 0.35% of these patients carry rare CNVs in the chromosomal locus 7q36.3. These rare copy number gains, or microduplications, were much less frequent (0.03%) among 7,431 healthy controls. It is particularly intriguing that all of the microduplications overlapped with — or occurred within 89 kilobases upstream of — a single gene, *VIPR2*.

The *VIPR2* gene encodes the VPAC2 receptor, which binds a brain chemical known as vasoactive intestinal peptide (VIP). The VIP–VPAC2 signalling pathway has been implicated in several neural and behavioural processes<sup>6</sup>, and is known to influence circadian rhythms<sup>7</sup> and cognition<sup>8</sup>. Vacic and co-workers' identification of several duplications in the protein-coding sequences at or near *VIPR2* suggests that alteration in VIP signalling is also a risk factor for schizophrenia.

The activity of the VPAC2 receptor is positively coupled with that of the intracellular enzyme adenylyl cyclase. And stimulation of this receptor increases the intracellular levels of the signalling molecule cyclic AMP, which is generated by adenylyl cyclase. To determine whether the *VIPR2*-associated CNVs have a physiological role, the authors<sup>2</sup> examined immune cells called lymphocytes obtained from subjects who had 7q36.3 microduplications and from control subjects. Specifically, they measured cAMP levels in response to the addition of VIP or another selective VPAC2 activator. Lymphocytes from patients carrying the microduplications showed significantly higher levels of cAMP than did cells from controls.

The implication of cAMP-mediated signalling in schizophrenia is particularly noteworthy because this molecule has several crucial functions, including effects on neuronal activity and gene transcription<sup>8</sup>. Changes in VPAC2 activity are one route by which this signalling pathway might be affected in schizophrenia. Unfortunately, how 7q36.3 microduplications affect the structure of VPAC2 and its coupling to adenylyl cyclase remains unknown.

VPAC2 is expressed throughout the nervous system, and so it is essential to identify where in the brain of patients with schizophrenia the microduplications in *VIPR2* are active and how they alter neuronal behaviour. Candidate sites include the amygdala, hippocampus and suprachiasmatic nuclei, which are respectively involved in emotion, learning and memory, and circadian rhythms. Interestingly, mice with targeted disruptions in VIP–VPAC2 signalling show alterations in neural processes affected in schizophrenia, including cognition<sup>9</sup>, responses to sensory stimuli<sup>10</sup> and sleep<sup>11</sup>.

Vacic and colleagues' study<sup>2</sup> coincides with another investigation<sup>12</sup> that also points to CNVs at 7q36.3 as potential risk factors for schizophrenia. Thus, converging evidence supports voluntary testing for CNVs at this locus to detect the likelihood of a person developing schizophrenia. Of course, such an endeavour is fraught with concerns over whether a person really wants to know that they are at risk of developing schizophrenia, or indeed over how this CNV influences particular symptoms associated with the disorder.

These studies<sup>2,12</sup> also identify VPAC2 as a potential target for therapy. The distribution of this receptor in mammalian tissues is reasonably well mapped, and VPAC2-selective compounds have been, and continue to be, developed. Potentially, therefore, drugs could be tailored to regulate VIP–VPAC2 signalling in distinct regions of the brain. A possible, though surmountable, problem is that neuropeptides other than VIP bind to VPAC2, and that VIP can signal through other receptors. Another cause for caution is that the relative incidence of CNVs at the 7q36.3 locus — although significant — is low and may account for only a small proportion of the risk. It is more likely, therefore, that VIP inhibitors could form part of a cocktail of drugs to treat the symptoms associated with schizophrenia. ■

**Hugh D. Piggins** is in the Faculty of Life Sciences, University of Manchester, Manchester M13 9PT, UK.  
e-mail: hugh.d.piggins@manchester.ac.uk

1. van Os, J. & Kapur, S. *Lancet* **374**, 635–645 (2009).
2. Vacic, V. *et al. Nature* **471**, 499–503 (2011).
3. Howes, O. D. & Kapur, S. *Schizophrenia Bull.* **35**, 549–562 (2009).
4. Tam, G. W. C., Redon, R., Carter, N. P. & Grant, S. G. N. *Biol. Psychiatry* **66**, 1005–1012 (2009).
5. Mulle, J. G. *et al. Am. J. Hum. Genet.* **87**, 229–236 (2010).



6. Moody, T. W. *et al.* *Curr. Opin. Endocrinol. Diabetes Obesity* **18**, 61–67 (2011).
7. Brown, T. M., Hughes, A. T. & Piggins, H. D. *J. Neurosci.* **25**, 11155–11164 (2005).
8. Pierre, S., Eschenhagen, T., Geisslinger, G. & Scholich, K. *Nature Rev. Drug Discov.* **8**, 321–335 (2009).
9. Chaudhury, D. *et al.* *BMC Neurosci.* **9**, 63 (2008).
10. Hughes, A. T., Fahey, B., Cutler, D. J., Coogan, A. N. & Piggins, H. D. *J. Neurosci.* **24**, 3522–3526 (2004).
11. Sheward, W. J. *et al.* *PLoS ONE* **5**, e9783 (2010).
12. Levinson, D. F. *et al.* *Am. J. Psychiatry* **168**, 302–316 (2011).

## GLOBAL WARMING

# The soot factor

**The surface warming due to emissions of black-carbon aerosols over the second half of the twentieth century has been identified in observations. These findings will inform debate over the climatic effects of controlling such emissions.**

JOHANNES QUAA

Anthropogenic emissions affect global climate. In a paper published in *Atmospheric Chemistry and Physics*, Jones *et al.*<sup>1</sup> add to understanding of this ‘forcing’ of climate in a study of one of the factors involved — black carbon, or soot.

Greenhouse gases, most prominently carbon dioxide, warm the planet by trapping infrared radiation. A less-recognized, but important, climate forcing is that caused by aerosols, small particles emitted during the combustion of fossil fuel and biofuel such as wood<sup>2</sup>. The additional anthropogenic aerosols reflect sunlight, and also act as cloud-condensation nuclei — particles on which water vapour can condense and form clouds, which reflect more sunlight if they are formed from greater numbers of droplets. Both processes cool the Earth’s surface.

Black carbon, however, is a special aerosol because it also strongly absorbs sunlight (Fig. 1) — precisely the reason it looks black. The immediate effect of this absorption of sunlight is to warm the atmosphere and cool the surface. But because the Earth system as a whole

absorbs sunlight rather than reflecting it, the overall result is a warming. Clouds may serve as a medium that strongly enhances this effect: if the atmosphere is heated as a result of the absorption of radiation, clouds may dissolve, which would probably constitute a further net warming. Aerosols quickly settle through gravity and are washed out by precipitation. Their typical lifetime in the lower atmosphere is thus relatively short (up to one week), so they are concentrated near their sources. The pattern of the forcing by anthropogenic aerosols thus reflects their main source regions.

Jones *et al.*<sup>1</sup> now identify the pattern of surface warming produced by anthropogenic black carbon in the observational record for the second half of the twentieth century. The analysis involved performing a set of climate-change simulations for the twentieth century, using a coupled atmosphere–ocean general circulation model and considering various combinations of individual forcings (greenhouse gases, aerosols including black carbon, ozone and land-use changes, as well as natural forcings such as volcanic emissions and changes in solar irradiation).

The authors split the temperature responses

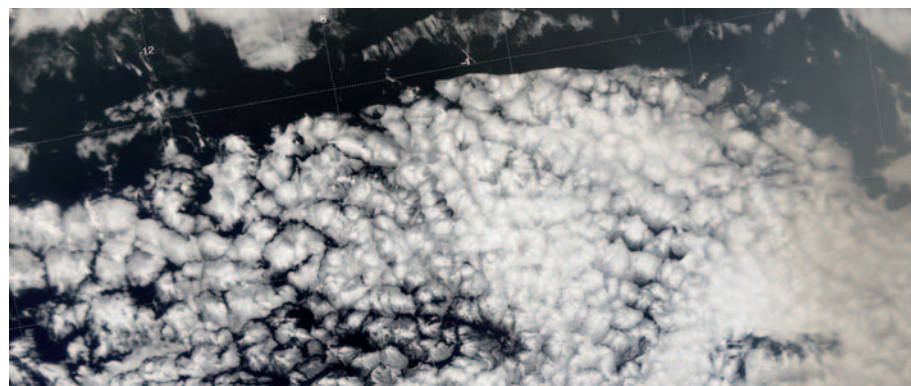
as estimated using these simulations into four groups — temperature changes due to natural forcings; changes due to anthropogenic greenhouse gases; changes due to other anthropogenic factors (aerosols other than black carbon, ozone, and land-use alteration); and changes due to black carbon. The ‘detection and attribution’ method to identify a temperature signal due to a specific forcing was then applied to try to discern the corresponding simulated patterns of temperature change in the observational record.

Focusing on the second half of the twentieth century (1950–99), Jones *et al.* find that natural forcing did not contribute significantly to the temperature trends, whereas all three groups of anthropogenic forcings did contribute significantly. This was previously known for the warming by greenhouse gases and the relative cooling by aerosols. But Jones *et al.* for the first time show that black carbon has warmed Earth by a statistically significant amount.

Their study<sup>1</sup> is particularly timely given the recent revival of discussion about whether control of black-carbon emissions might help to mitigate global warming. The scientific debate on this topic focuses on the net effect of a reduction in anthropogenic black carbon because, in addition to the warming caused by absorption of sunlight, black carbon may affect climate by other means. When mixed with other aerosols, black-carbon aerosols may serve as cloud-condensation nuclei, altering cloud reflectivity and probably having a cooling influence. By contrast, when deposited on snow, black carbon may act as a warming agent. The amount of sunlight reflected is reduced if a snow surface polluted by soot is less bright, an effect that may become stronger if the snow is heated so much that it melts.

These two contrasting effects of black carbon were not included in Jones and colleagues’ model. However, if either the cooling effect by enhanced cloud reflectivity was large enough to compensate for the warming due to the absorption of sunlight plus potential dissolution of clouds, or the warming by pollution of snow was greater than the other effects, the simulated pattern of temperature change probably would not have been detectable in the observations. Thus, this study is also a notable contribution to the debate about which of black carbon’s impacts on climate is most important.

In quantitative terms, considerable uncertainty remains. In Jones and colleagues’ study, this is most manifest in the fact that a substantial warming of 0.21 kelvin due to black carbon was detected for 1950–99, whereas it is only 0.14 kelvin for a slightly shifted period, 1957–2006. The likely reasons are that this second period was more strongly affected by the 1991 eruption of Mount Pinatubo, and that the relative importance of black carbon diminished after the late 1990s in many regions. It is also noteworthy that other anthropogenic climate forcings had a much stronger impact on the global



**Figure 1 | Plume of soot aerosol overlying clouds above the eastern Atlantic Ocean.** This image, taken by NASA’s MODIS satellite instrument, shows a black-carbon plume (lower right) that emanated from biomass burning in west-central Africa. The aerosol absorbs radiation and so reduces the radiation being reflected back into space from the underlying clouds. Jones *et al.*<sup>1</sup> show that anthropogenic emissions of black-carbon aerosols have warmed Earth’s climate to a statistically significant level over the period 1950–99. (Faint vertical lines are lines of longitude; the image is about 2,200 kilometres across.)

temperature — Jones *et al.* find a three to four times greater warming by greenhouse gases. In the context of mitigation of climate change, a further consideration is that black carbon is usually co-emitted with other aerosols, such as sulphate, that cool the climate. Reductions in total aerosol output might be desirable for public-health purposes. In the context of climate change, however, appropriate technical means would have to be applied to reduce the warming influence of black carbon but not the (probably larger) cooling effect of other aerosols.

Finally, it is also necessary to remember

that anthropogenic aerosols, including black carbon, have a very short atmospheric lifetime compared with that of greenhouse gases. The gases typically have lifetimes of centuries and longer<sup>3</sup>, compared with days for aerosols. This implies that, if implementation of emission-reduction strategies were indeed to be feasible, climate-change mitigation by cutting black-carbon emissions could be effective fast. But it also suggests that the relative importance of black carbon will in any case gradually diminish, given that greenhouse gases are long-lived and that they will continue to accumulate

in the atmosphere as long as anthropogenic emissions of these gases continue. ■

Johannes Quaas is at the Max Planck Institute for Meteorology, D-20146 Hamburg, Germany.

e-mail: johannes.quaas@zmaw.de

1. Jones, G. S., Christidis, N. & Stott, P. A. *Atmos. Chem. Phys.* **11**, 799–816 (2011).
2. Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
3. Solomon, S. *et al.* *Proc. Natl Acad. Sci. USA* **106**, 1704–1709 (2009).

## DEVELOPMENTAL BIOLOGY

# A mouse is not a cow

**Early cell-lineage decisions during embryonic development differ between mice and cows. This finding calls for a re-examination of developmental variations across mammals, but does not undermine use of the mouse as a model organism.**

JANET ROSSANT

The mammalian blastocyst is a thing of beauty. Over a period of a few days after the union of an egg with sperm, the fertilized egg divides to generate this tiny hollow sphere of cells, which has a cluster of enclosed cells at one end of the fluid-filled cavity. The outer cells are called the trophectoderm and the inner cells are, inventively, named the inner cell mass. But when do cells commit to becoming one or the other, and how? Writing in *Developmental Cell*, Berg *et al.*<sup>1</sup> show that the answers to these questions are not the same for mice and cattle.

Pluripotency — a cell's ability to differentiate into all cell types of the body — is a common property of the inner cell mass (ICM) of all mammalian blastocysts and is always associated with the expression and function of the transcription factor Oct4. The trophectoderm, which later generates all of the specialized layers of the placenta, also expresses a number of lineage-restricted transcription factors, most notably Cdx2.

In mice, deletion of either the *Oct4* gene (also known as *Pou5f1*) or the *Cdx2* gene leads to the formation of abnormal blastocysts: ICM cells of *Oct4*-mutant blastocysts express trophectoderm markers and lose pluripotency<sup>2</sup>, whereas the outer cells of *Cdx2*-mutant blastocysts express pluripotency markers such as Oct4 ectopically and fail to differentiate

further down the trophectoderm lineage<sup>3</sup>. This suggests a model — albeit an overly simplistic one — whereby restricted expression of Oct4 and Cdx2 leads to reciprocal repression of the opposing lineage and establishes cell fate.

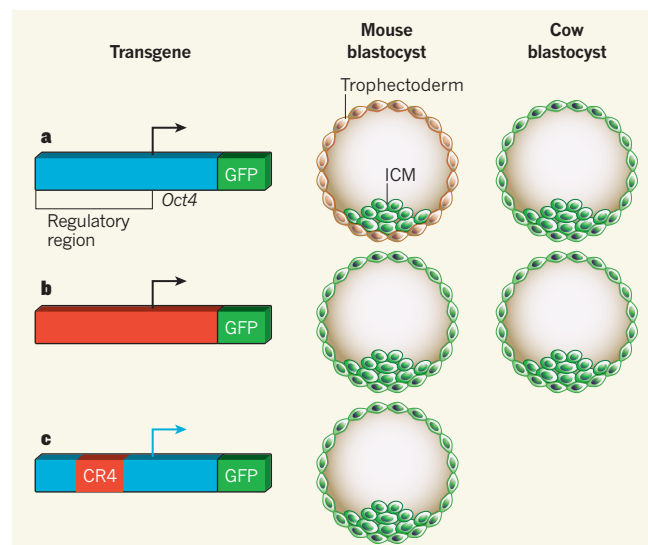
Berg *et al.*<sup>1</sup> asked whether this model applies to cell-fate decisions in cows. They find that, unlike in mice, Oct4 expression is not restricted only to the ICM during the early stages of cow blastocyst development. Instead, Oct4 is co-expressed with Cdx2 in the trophectoderm for some time after the beginning of blastocyst

formation. This observation is consistent with previous reports and has also been made for pig and human embryos (for example, see refs 4, 5). Even in the mouse, Oct4 expression overlaps with Cdx2 expression during the late cleavage and early blastocyst stages of embryonic development, and is restricted to the ICM only by the fully expanded blastocyst stage<sup>3</sup>.

So why is Oct4 expression maintained for longer in the cow trophectoderm than in its mouse equivalent? Through experiments involving cow blastocysts engineered to express a fluorescently tagged version of mouse *Oct4* (the mouse *Oct4-GFP* transgene), Berg and co-workers show that the factors that restrict Oct4 expression to the ICM are not available, or not functional, in the cow blastocyst (Fig. 1a). Cdx2 could be one such factor, but the authors' data suggest that this protein has a role only later during cow embryonic development. However, Berg and colleagues do not investigate whether the role of Cdx2 in restricting Oct4 expression is simply delayed in the cow embryo, nor whether Oct4 is ectopically expressed later during development in embryos treated to express reduced levels of Cdx2.

The paper<sup>1</sup> shows that a mouse *Oct4-GFP* transgene containing the bovine Oct4 regulatory elements is expressed in both the ICM and trophectoderm in fully expanded blastocysts of both the cow and the mouse (Fig. 1b). This suggests that Cdx2, which is active in mouse blastocysts, is not the only factor that affects the timing of Oct4 repression. It also indicates that bovine regulatory elements do not respond to the factors that downregulate Oct4 in mouse blastocysts.

Of the four evolutionarily conserved regulatory regions around the Oct4 locus, CR4 shows the most sequence divergence between the mouse and the cow. When Berg *et al.* replaced mouse CR4 with the cow version in the mouse *Oct4-GFP* construct, it behaved like the cow gene in the mouse blastocysts (Fig. 1c). Thus changes in both DNA



**Figure 1 | Oct4 regulation in mouse and cow blastocysts.** a, Berg *et al.*<sup>1</sup> find that the expression of a GFP fluorescent reporter transgene controlled by the regulatory elements of the mouse *Oct4* gene (blue) is restricted to the inner cell mass (ICM) in mouse blastocysts but not cow blastocysts. b, The same transgene, but containing the bovine Oct4 regulatory elements (red), is not restricted to the ICM in either cow or mouse blastocysts. c, The authors narrow down this effect of bovine regulatory elements to the CR4 region.



regulatory regions and the factors that bind to such sequences drive differences in the regulation of Oct4 expression between mouse and cow blastocysts.

It would be interesting to test, in transgenic mice, whether regulatory elements of the human *OCT4* gene behave like the mouse or the cow sequences. Although human blastocysts, like those of domestic animals, express Oct4 in the trophectoderm for an extended period compared with mice, the period of overlap of Cdx2 and Oct4 expression is only slightly longer than in the mouse. Human OCT4 is clearly restricted to the ICM by day 6 before embryo implantation<sup>6</sup>.

But why do these regulatory differences exist among the blastocysts of different mammals? Evolutionarily, the placenta is a recent invention, and still seems to be a work in progress. There is huge variation in trophectoderm and placental morphology across different mammalian species, accompanied by recent evolutionary divergence in placenta-specific gene families<sup>7</sup>. For example, a mouse blastocyst attaches and implants in the uterus by embryonic day 5 (E5); a human blastocyst grows a little larger but then implants by E7–9 with highly invasive trophoblast outgrowth; and in cows, pigs and sheep the blastocyst floats in the uterus for 2–3 weeks before attaching.

Berg *et al.* propose that such differences lead to earlier restriction of trophectoderm cell fate in the mouse than in the cow. Indeed, results of their experiments — involving chimaeric blastocysts generated by mixing trophectoderm cells from different stages of development with host embryos — support this proposal.

In a remarkable technical tour de force, they also transferred the chimaeric cow blastocysts to recipient cows and recovered them later in development to show that early trophectoderm cells can contribute to developing ICM derivatives. This is one of the first attempts to test the timing of lineage restriction in a species other than the mouse.

This study emphasizes the need to explore the timing and mechanism of functional lineage restriction in blastocysts of different mammals, including humans. Differences in these parameters may underlie the known difficulty in deriving validated pluripotent embryonic stem cells and trophoblast stem cells from many mammalian species. Although fibroblasts have been reprogrammed into induced pluripotent stem cells in several domestic species, including the cow, these lines often depend on continued expression of exogenous reprogramming factors. Clearly, we need a better understanding of the control of pluripotency in all these species.

As we learn more about the precise details of mouse blastocyst development, we must be constantly evaluating similarities and differences between them and those of humans and other species. This will help us to truly understand mammalian embryo diversity. ■

Janet Rossant is in the Program in Developmental and Stem Cell Biology, Hospital for Sick Children, and in the Department of Molecular Genetics, University of Toronto, Ontario M5G 1X8, Canada.  
e-mail: janet.rossant@sickkids.ca

## MOLECULAR BIOLOGY

# A fly in the face of genomics

The modENCODE project uses integrative analysis to annotate genomic elements in the fruitfly and a nematode worm. The first fly data have now been published. [SEE ARTICLES P.473 & P.480](#) & [LETTER P.527](#)

EILEEN E. M. FURLONG

The fruitfly *Drosophila melanogaster* is an exceptional model for dissecting the basic principles of biology, development and disease. It is amenable to genetic manipulation using tools developed over more than a century; and its genome shares extensive genetic content with humans. The first draft of the *Drosophila* genome was released a decade ago<sup>1</sup>, and with subsequent updates its annotation is in a 'mature' state. Nevertheless, more than half of the predicted genes have been awaiting experimental verification of their structure — the location of promoter sequences, of boundaries of protein-coding and non-coding sequences, and of transcription termini. The modENCODE consortium project aims to address this issue and to identify new genes and genomic elements in the fly genome<sup>2</sup>. Here I focus on the first wave of papers, including three in this issue<sup>3–5</sup>, which describes the fly data so far.

To determine which genes are expressed at specific stages of development, Graveley *et al.*<sup>3</sup> (page 473) generated high-resolution expression data, which are complemented by an analysis of 25 *Drosophila* cell lines<sup>6,7</sup>. These efforts identified almost 2,000 new genes that encode proteins or non-coding RNAs. They also extensively refine existing annotation by describing more than 3,000 new promoter sequences<sup>7</sup>, roughly 53,000 new or revised exon sequences<sup>3</sup>, a threefold increase in RNA-splicing events<sup>3</sup> and a tenfold increase in RNA-editing events<sup>3</sup>. Notably, most of the RNA-editing and -splicing events occur at precise stages of the *Drosophila* life cycle, indicating extensive temporal regulation of these post-transcriptional events by as-yet poorly understood mechanisms. This comprehensive view of the fly transcriptome<sup>3,6,7</sup> reveals that some 75% of the organism's genome is transcribed at

1. Berg, D. K. *et al.* *Dev. Cell* **20**, 244–255 (2011).
2. Nichols, J. *Cell* **95**, 379–391 (1998).
3. Strumpf, D. *et al.* *Development* **132**, 2093–2102 (2005).
4. Rossant, J. *Reprod. Fertil. Dev.* **19**, 111–118 (2007).
5. Blomberg, L., Hashizume, K. & Viebahn, C. *Reproduction* **135**, 181–195 (2008).
6. Chen, A. E. *et al.* *Cell Stem Cell* **4**, 103–106 (2009).
7. Wildman, D. E. *Placenta* **32**, 142–145 (2011).

one stage or another — in line with the widespread transcription observed in other species.

Post-translational histone modifications covering a gene's promoter or coding region provide telltale signatures of the expression status of a gene and thereby present another way to identify functional elements in the genome. Two of the modENCODE studies involved mapping such chromatin marks in *Drosophila* cell lines<sup>4</sup> and at 11 stages of its life cycle<sup>5</sup>.

By examining the distribution of 18 histone modifications in two cell lines, Kharchenko *et al.*<sup>4</sup> (page 480) identified nine prominent chromatin signatures, which complement those defined previously<sup>8</sup>. Clues to their function come from information on chromatin accessibility and transcriptional activity, revealing chromatin signatures that distinguish between active and inactive genes, active promoters, and the location of new putative regulatory elements. The authors' global analyses<sup>4</sup> extend previous studies<sup>9–12</sup> indicating that the Polycomb system — a group of chromatin-binding proteins traditionally associated with stable, long-term gene repression during embryonic development — can also function dynamically and associate with promoters that are actively transcribed or seem poised for activation.

Deposition of chromatin marks is linked to the enzymatic activity of RNA polymerase during the initiation and elongation steps of transcription; this activity is regulated by transcription factors bound to *cis*-regulatory elements — proximal and distal sequences that affect gene expression. To understand how transcription is regulated, Nègre *et al.*<sup>5</sup> (page 527) made a systematic effort to identify all *cis*-regulatory elements by examining the occupancy of 38 transcription factors and other chromatin-regulatory proteins at different stages of development. The result is a collection of around 20,000 putative regulatory elements that include insulators, enhancers and

## COSMOLOGY

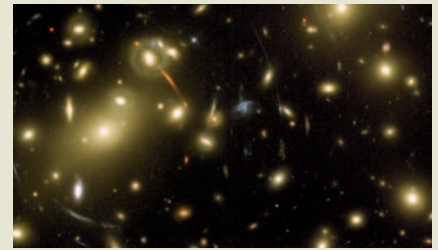
# Lenses under the lens

If you were to pop into a cosmology conference today, the chances are that you would see this image in at least one presentation. It is a striking snapshot of a cluster of galaxies acting as a gravitational lens: the cluster bends light from galaxies lying behind it and ‘smears’ the light to produce multiple images and giant arcs.

As pretty as their effects are, gravitational lenses are giving cosmologists a few headaches. For example, the observed incidence of giant arcs and their distance from the clusters’ centres, which marks the size of features called Einstein rings, indicate that these clusters may have a

stronger ‘lensing’ ability than expected in the framework of the currently accepted model of the cosmos. In a paper to appear in *Astronomy & Astrophysics*, Meneghetti *et al.* describe an analysis that advances our understanding of these systems (M. Meneghetti *et al.* Preprint at <http://arXiv.org/abs/1103.0044>; 2011).

The authors compared the lensing ability of a numerically simulated sample of clusters with that of a sample of well-characterized, X-ray-luminous clusters obtained by the MAssive Cluster Survey (MACS). In contrast to earlier studies, their simulations factor in elements known to affect lensing power — for example, the fact that the



lenses are complex three-dimensional structures. They found that the simulated clusters produce 50% fewer arcs than do the observed MACS clusters, and that the median size of Einstein rings differs by 25% between the two samples. These are much smaller discrepancies between theory and observation than previously reported. But as the authors themselves concede, more data are needed to confirm their findings. **Ana Lopes**

promoters. Of the more than 2,000 putative promoters, 50% are already confirmed<sup>13</sup>. The locations of about 14,500 putative *cis*-regulatory elements were also identified. Unexpectedly, one class of active promoters does not contain the characteristic chromatin mark H3K4me3, suggesting that the genes they regulate use an alternative mode of transcriptional initiation.

Integrating the binding patterns of all transcription factors leads to hypotheses of transcription-factor partnerships, involving co-binding to regulatory elements<sup>5,7</sup>. But overlays of transcription-factor binding should be interpreted cautiously, particularly for factors with non-tissue-specific or partially overlapping expression: regions that are co-targeted by multiple factors are not necessarily co-bound in the same cells. Nevertheless, the complexity of some co-targeted regions is intriguing. The modENCODE researchers identified regions in the genomes of both *Drosophila*<sup>5,7</sup> and the nematode *Caenorhabditis elegans*<sup>14</sup> — the other model organism on which the project focuses — that are highly occupied by transcription factors. It remains to be determined what function, if any, such regions have in transcription.

This first phase of modENCODE has made a significant impact on refining the annotation of the *Drosophila* genome, which forms the foundation of a large body of research conducted in this organism. But where should the project go from here? First, there is the issue of completion. With the new data, the annotation of genes may be 80% complete, but the job is far from over. Despite the huge depth of coverage, almost 1,500 known genes could not be identified in any experiments<sup>4</sup>. Analysis of specific subpopulations of cells and tighter staging of the developmental process should greatly improve sensitivity.

Completing annotation of the ‘regulatory genome’ is much more challenging. Although

the location of putative enhancer elements can be identified, determining which of these regions are functional, and when, is a huge task. Understanding the regulation of enhancer activity requires knowledge of which transcription factors are binding to them, in which cell types, and when. Scaling this up to the roughly 700 predicted *Drosophila* transcription factors is a monumental undertaking, but feasible given current tagging technologies<sup>15,16</sup>.

A major drawback of the data sets is their lack of temporal and spatial resolution. Although cells in culture are extremely useful for identifying core properties of basic cellular processes, such immortalized cells, devoid of their developmental context, cannot substitute for cells within a developing embryo. On the other hand, whole-embryo studies provide merged signals from all cells in the embryo, giving no information on the tissue in which a gene, promoter or chromatin state is active. Many of the transcription factors examined are expressed across a broad range of tissues, which has the advantage of covering a wide range of *cis*-regulatory elements. But merged transcription-factor occupancy signals from multiple tissues make it very difficult to disentangle regulatory connections and thus to build reliable regulatory networks.

The general absence of functional information is perhaps the most serious limitation of the current work and a major challenge for all genomics projects. Such information is essential to understand the relevance of regulatory connections. Examining mutants was understandably beyond the scope of the present studies, but, moving forward, there is a clear need to integrate diverse types of functional data in order to make the transition from correlations to regulatory function. The thousands of *Drosophila* mutants available should provide a useful resource for this.

We can view this work<sup>3–5</sup> as an important

chapter in a long book. The data — all freely available<sup>17</sup> — provide an excellent resource for identifying putative genes and regulatory elements that might be active at a particular stage of development. The sheer volume of new transcripts and putative regulatory elements, and the inherent complexity of their interactions, demonstrates how far the project has come, but also highlights the challenges that lie ahead to convert this wealth of information into regulatory networks that describe the transformation of a fertilized egg into a complex multicellular organism. To reach this goal, researchers must integrate new types of experiments that will address the function of, and connections between, genomic regions at high spatio-temporal resolution. With this in mind, we can envisage a next phase of exciting studies that will tackle these issues, and so look forward to seeing what comes next. ■

**Eileen E. M. Furlong** is at the *Genome Biology Unit, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany.* e-mail: [furlong@embl.de](mailto:furlong@embl.de)

1. Adams, M. D. *et al.* *Science* **287**, 2185–2195 (2000).
2. Celniker, S. E. *et al.* *Nature* **459**, 927–930 (2009).
3. Graveley, B. R. *et al.* *Nature* **471**, 473–479 (2011).
4. Kharchenko, P. V. *et al.* *Nature* **471**, 480–485 (2011).
5. Nègre, N. *et al.* *Nature* **471**, 527–531 (2011).
6. Chervas, L. *et al.* *Genome Res.* **21**, 301–314 (2011).
7. The modENCODE Consortium *Science* **330**, 1787–1797 (2010).
8. Filion, G. J. *et al.* *Cell* **143**, 212–224 (2010).
9. Papp, B. & Müller, J. *Genes Dev.* **20**, 2041–2054 (2006).
10. Kwong, C. *et al.* *PLoS Genet.* **4**, e1000178 (2008).
11. Oktaba, K. *et al.* *Dev. Cell* **15**, 877–889 (2008).
12. Enderle, D. *et al.* *Genome Res.* **21**, 216–226 (2011).
13. Hoskins, R. A. *et al.* *Genome Res.* **21**, 182–192 (2011).
14. Gerstein, M. B. *et al.* *Science* **330**, 1775–1787 (2010).
15. Venken, K. J. T. *et al.* *Nature Methods* **6**, 431–434 (2009).
16. Ejsmont, R. K., Sarov, M., Winkler, S., Lipinski, K. A. & Tomancak, P. *Nature Methods* **6**, 435–437 (2009).
17. [www.modencode.org](http://www.modencode.org)



# Catalytic Z-selective olefin cross-metathesis for natural product synthesis

Simon J. Meek<sup>1</sup>, Robert V. O'Brien<sup>1</sup>, Josep Llaveria<sup>1</sup>, Richard R. Schrock<sup>2</sup> & Amir H. Hoveyda<sup>1</sup>

Alkenes are found in many biologically active molecules, and there are a large number of chemical transformations in which alkenes act as the reactants or products (or both) of the reaction. Many alkenes exist as either the *E* or the higher-energy *Z* stereoisomer. Catalytic procedures for the stereoselective formation of alkenes are valuable, yet methods enabling the synthesis of 1,2-disubstituted *Z* alkenes are scarce. Here we report catalytic *Z*-selective cross-metathesis reactions of terminal enol ethers, which have not been reported previously, and of allylic amides, used until now only in *E*-selective processes. The corresponding disubstituted alkenes are formed in up to >98% *Z* selectivity and 97% yield. These transformations, promoted by catalysts that contain the highly abundant and inexpensive metal molybdenum, are amenable to gram-scale operations. Use of reduced pressure is introduced as a simple and effective strategy for achieving high stereoselectivity. The utility of this method is demonstrated by its use in syntheses of an anti-oxidant plasmalogen phospholipid, found in electrically active tissues and implicated in Alzheimer's disease, and the potent immunostimulant KRN7000.

Carbon-carbon double bonds reside within a large variety of molecules that possess desirable properties<sup>1</sup>, and catalytic cross-metathesis<sup>2</sup> (CM; Fig. 1) represents one of the most attractive approaches to stereoselective preparation of these versatile functional groups. Through fusion of two terminal alkenes, available in ample quantities as by-products of petroleum purification or readily accessed by a variety of methods, 1,2-disubstituted alkenes can be obtained; the other product generated is gaseous ethylene. However, the only reported instances of *Z*-selective CM (65–90% *Z*) involve substrates with an *sp*-hybridized substituent (acrylonitrile or enynes<sup>3,4</sup>). In an efficient *Z*-selective CM, it is not only required that reaction between the two substrates proceed selectively (versus homocoupling), it must exhibit a preference for the thermodynamically less favoured stereoisomer (Fig. 1). The inherent reversibility of olefin metathesis (products can re-enter the catalytic cycle) and the higher reactivity of *Z* alkenes (versus *E* isomers) further exacerbate the problem. Through careful consideration of various mechanistic aspects of the process, conditions must be identified where the catalyst promotes CM but fails to react with the product *Z* alkene to effect equilibration, favouring the lower energy *E* isomer.

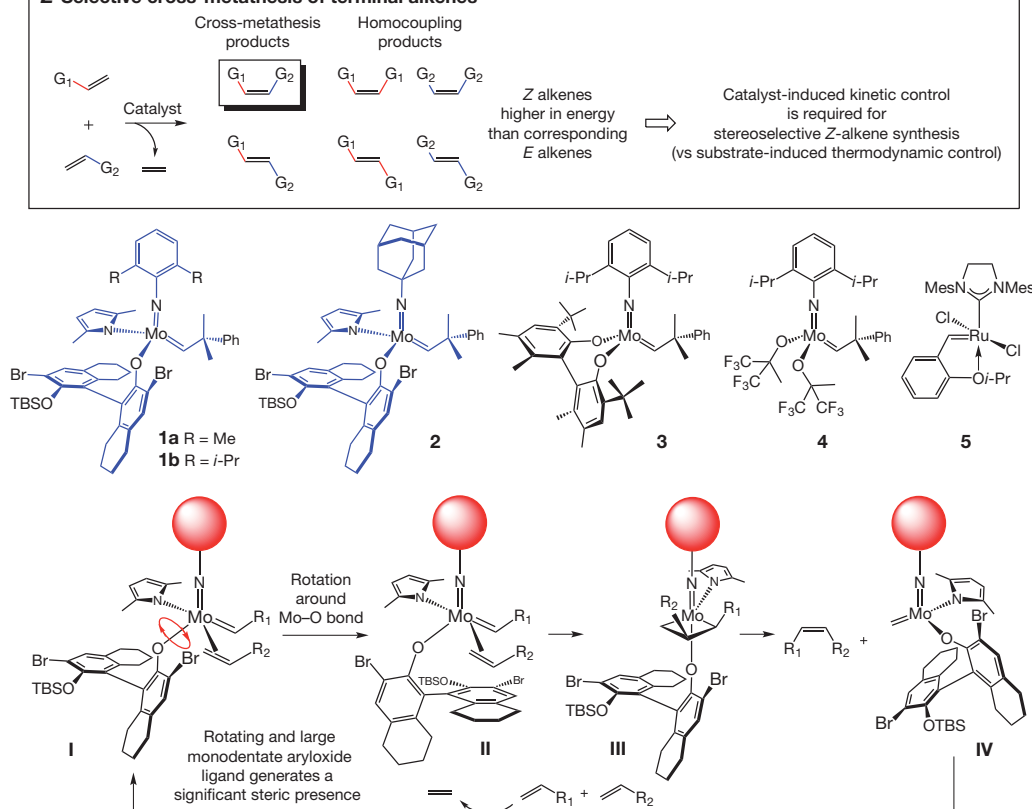
## Challenges of catalytic *Z*-selective cross-metathesis

As the preliminary steps towards the eventual development of an efficient class of *Z*-selective CM reactions, we investigated two related—but much simpler—versions of the process. Alkylidenes and carbenes **1**–**5** (Fig. 1) represent the catalyst classes used in our studies. Stereogenic-at-Mo **1a**, **1b** and **2**<sup>5,6</sup> were recently designed in these laboratories to promote enantioselective ring-closing metathesis; theoretical<sup>7</sup> and experimental explorations<sup>8</sup> suggest that these complexes exhibit high activity partly as the result of stereoelectronic effects induced by the electron donor pyrrolide and acceptor monoaryloxide ligands. The fluxional nature of complexes such as **1a**, **1b** and **2**, facilitated by the absence of rigid bidentate ligands, allows the metal alkylidenes to adapt to the structural strains imposed during the

catalytic cycle. As a result, the stereogenic-at-Mo complexes are generally more effective olefin metathesis catalysts than other Mo-based complexes **3**<sup>9</sup> and **4**<sup>10</sup> or Ru carbene **5**<sup>11</sup>. We thus established that alkylidene **2** readily catalyses *Z*-selective alkene formation through ring-opening/cross-metathesis (ROCM)<sup>12</sup> with strained oxabicyclic alkenes and styrenes. Homocoupling of terminal alkenes was subsequently shown to proceed with high efficiency and *Z* selectivity in the presence of members of the same catalyst class<sup>13</sup>. The general mechanistic features that engender *Z* selectivity in the above reactions, and would be expected to do so in a CM process, are depicted in Fig. 1. The preference for *Z* alkene formation can be attributed to the ability of the large monodentate aryloxide to rotate freely (compare **1** in Fig. 1), causing the incoming alkene to be oriented such that its substituent (*R*<sub>2</sub>) is situated *syn* to that of the alkylidene (*R*<sub>1</sub>).

Designing a *Z*-selective CM is substantially more difficult: in a homocoupling, only one type of alkene is involved and no more than two stereoisomeric alkenes can be formed; in contrast, there are two substrates in CM, which can generate up to six different products. In the case of a catalytic ROCM<sup>14</sup>, a strained cyclic alkene and a terminal alkene that are reluctant to undergo homocoupling (for example, a styrene) must be selected as substrates for the catalytic process to be efficient<sup>15</sup>. Transformations are carefully crafted such that the alkylidene derived from the terminal alkene favours association with the cyclic alkene (versus another of the same type) in the ring-opening stage, generating a new Mo complex that strongly prefers to react with a sterically less demanding terminal alkene (CM stage). The possibility of a transformation between the alkylidene generated through ring-opening and another strained—but more hindered—cyclic alkene is thus discouraged (that is, minimal homocoupling or oligomerization). Such deliberate orchestration is not feasible with catalytic CM, where both alkenes are mono-substituted and, in contrast to ROCM, there is no relief of ring strain to be manipulated.

<sup>1</sup>Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02467, USA. <sup>2</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

**Z-Selective cross-metathesis of terminal alkenes**

**Figure 1** | A catalytic CM reaction can afford as many as six alkenes, so the challenge is designing an efficient process that favours formation of the cross products. Particularly difficult is the development of a process that affords the higher-energy Z alkene predominantly. To accomplish a Z-selective CM, a variety of catalysts were considered, such as stereogenic-at-Mo complexes (**1**, **2**) or other previously reported Mo- and Ru-based complexes

(**3**–**5**). The structural flexibility of the stereogenic-at-metal complexes **1** and **2** can give rise to exceptional reactivity, and free rotation around the Mo–O bond of these alkylidenes might serve as the basis for development of highly Z-selective olefin metathesis reactions of terminal alkenes. The sphere represents an appropriate size imido substituent.

**Z-Selective cross-metathesis of enol ethers**

We began by evaluating the ability of stereogenic-at-Mo complexes to promote transformations of enol ethers, a class of substrates for which a CM reaction has not been previously reported (*E*- or *Z*-selective); the resulting products have proven to be of utility in chemical synthesis and can be found in biologically active molecules (see below). In the presence of 2.5 mol% **1a**, CM between **6** and **7** (entry 1, Table 1) proceeds to 85% conversion to afford disubstituted enol ether **8a** in 98% *Z* selectivity and 73% yield. With **1b**, which bears a more sizeable 2,6-di-*i*-propyl-arylimido unit, the reaction is completely *Z*-selective (>98% *Z*) but 47% conversion is achieved within the same time span. When alkylidene **2** is used, CM proceeds to 37% conversion and

>98% *Z*-**8a** is generated; further transformation is not observed after six hours. Mo-based diolate **3** and Ru carbene **5** do not promote CM, and achiral Mo complex **4** catalyses a non-selective transformation (47.5% *Z*). Thus, stereogenic-at-Mo complexes prove to be effective in promoting enol ether CM, and although **1b** or the less hindered **2** also afford exceptional stereoselectivity, neither delivers the efficiency of **1a**. The 2,6-dimethylphenylimido **1a** therefore offers the best balance between activity and stereoselectivity. Such performance variations may be observed because catalyst turnover is slower with the more sizeable **1b** whereas the methylidene of the relatively unhindered **2** (compare **IV**, Fig. 1) might suffer from a shorter life span. Consistent with the above scheme, 82% **8a** is formed when CM with **1b** is allowed to continue for 16 hours; in contrast, conversion with **2** after 10 minutes or two hours is nearly identical (~38%).

There are several, mechanistically revealing, reasons for use of excess enol ether. CM generates a Mo-methylidene; this unhindered alkylidene can readily react with the *Z*-alkene product, reverse CM, cause equilibration and lower stereoselectivity. An enol ether reacts with a methylidene complex, circumventing diminution in *Z* selectivity. The more stable alkoxy-substituted alkylidene, generated from reaction of a methylidene complex and an enol ether (**I** in Fig. 1 with  $R_1 = \text{On-Bu}$ ), can undergo productive CM, giving rise to longer catalyst lifetime and improved turnover numbers. Furthermore, generation of the aforementioned alkoxy- or aryloxy-containing alkylidene means less of the alkyl-substituted derivative is formed and homocoupling of the aliphatic alkene is minimized. Owing to electronic factors, productive reaction between an enol ether-derived alkylidene and another O-substituted alkene is disfavored<sup>2</sup>. However, as use of excess enol ether is wasteful, we decided to examine the efficiency of the CM with

**Table 1** | Examination of various catalysts for CM with an enol ether

| $\text{On-Bu} + \text{Ph-CH=CH}_2 \xrightarrow[\text{C}_6\text{H}_6, 22^\circ\text{C}]{2.5 \text{ mol\% Mo or Ru complex}} \text{Bn-CH=CH-On-Bu}$ |           |        |                        |                        |                                  |
|---|-----------|--------|------------------------|------------------------|----------------------------------|
| Entry no.   | Complex   | Time   | Conv. (%) <sup>*</sup> | Yield (%) <sup>†</sup> | <i>Z</i> : <i>E</i> <sup>*</sup> |
| 1   | <b>1a</b> | 2 h    | 85                     | 73                     | 98:2                             |
| 2   | <b>1b</b> | 2 h    | 47                     | ND                     | >98:2                            |
| 3   | <b>2</b>  | 2 h    | 37                     | ND                     | >98:2                            |
| 4   | <b>3</b>  | 2 h    | <2                     | NA                     | NA                               |
| 5   | <b>4</b>  | 10 min | 80                     | ND                     | 47.5:52.5                        |
| 6   | <b>5</b>  | 24 h   | <2                     | NA                     | NA                               |

The reactions were carried out in purified benzene under an atmosphere of nitrogen gas; 10 equiv. of **6** was used (see Supplementary Information for details). NA, not available; ND, not determined.

<sup>\*</sup> Conversion (conv.) and *Z*:*E* ratios were measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures; the variance of values is estimated to be  $\leq \pm 2\%$ .

<sup>†</sup> Yield of isolated product after purification; the variance of values is estimated to be  $\leq \pm 5\%$ .

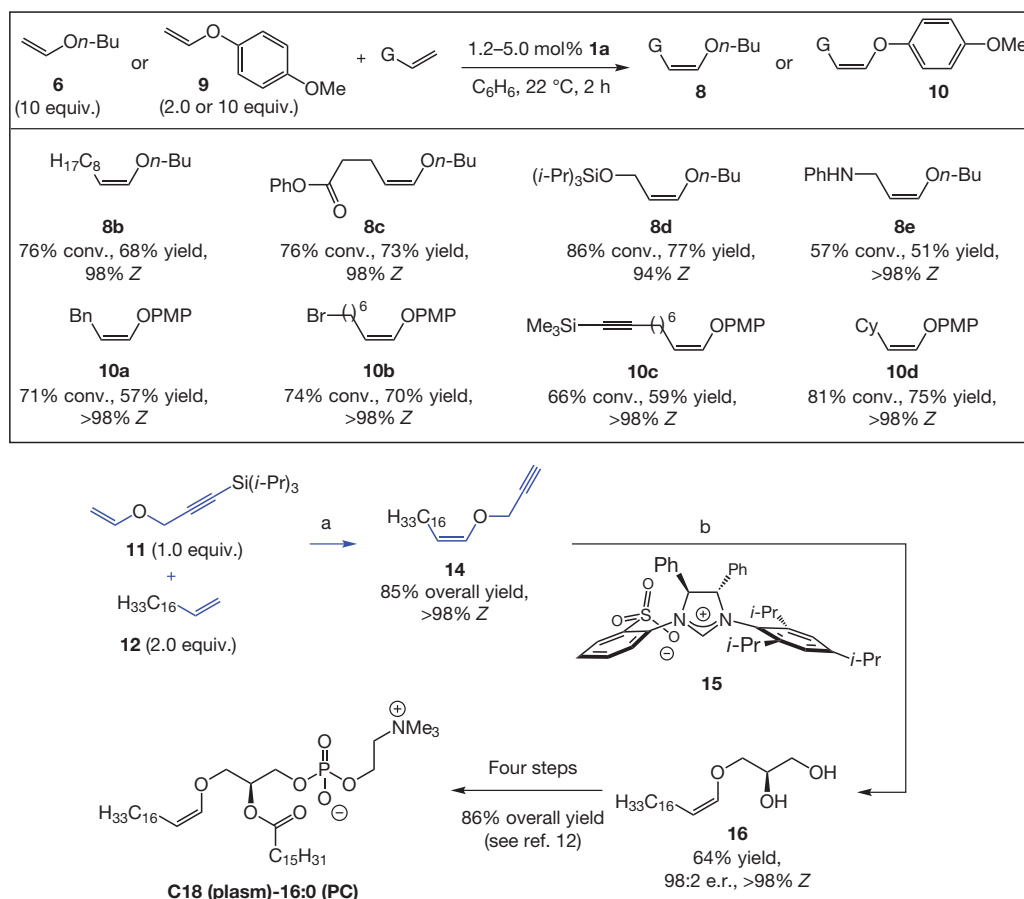


varying amounts of **6** (see Supplementary Information for details). The latter studies established that, although fewer equivalents of **6** lead to reduced *Z* selectivity and competitive homocoupling, with 5 equiv. of the inexpensive and commercially available enol ether, **8a** can be obtained in 93:7 *Z:E* selectivity and 71% yield (7% homocoupled product). Excess enol ether **6** does not complicate product isolation, as this inexpensive reagent is volatile and can be easily removed *in vacuo*.

*Z*-Disubstituted enol ethers are obtained in 57–77% yield through exceptionally stereoselective (94% to >98% *Z*) CM with Mo alkylidene **1a** (Fig. 2). Alkyl- (**8**) or aryl-substituted (**10**) *Z* enol ethers as well as those that bear a carboxylic ester (**8c**), a secondary amine (**8e**), a bromide (**10b**) or an alkyne (**10c**) are readily accessed. Reactions with the more electron-deficient enol ether **9** and the relatively electron-rich alkenes proceed with 2.0 equiv. of the aryl-substituted enol ether; in contrast, 10 equiv. of alkyl-substituted and easily removable **6** are required for similar efficiency. Such variations probably occur because when **9** is used there is a better electronic match<sup>3</sup> between the Mo-alkylidenes derived from the cross partners and either of the two alkenes, favouring CM versus homocoupling. Only 1.2 mol% **1a** and 2.0 equiv. of the *p*-methoxyphenylenol ether (for example, **10a**, **10b** and **10d**, Fig. 3c) are sufficient for an effective and exceptionally *Z*-selective CM to take place.

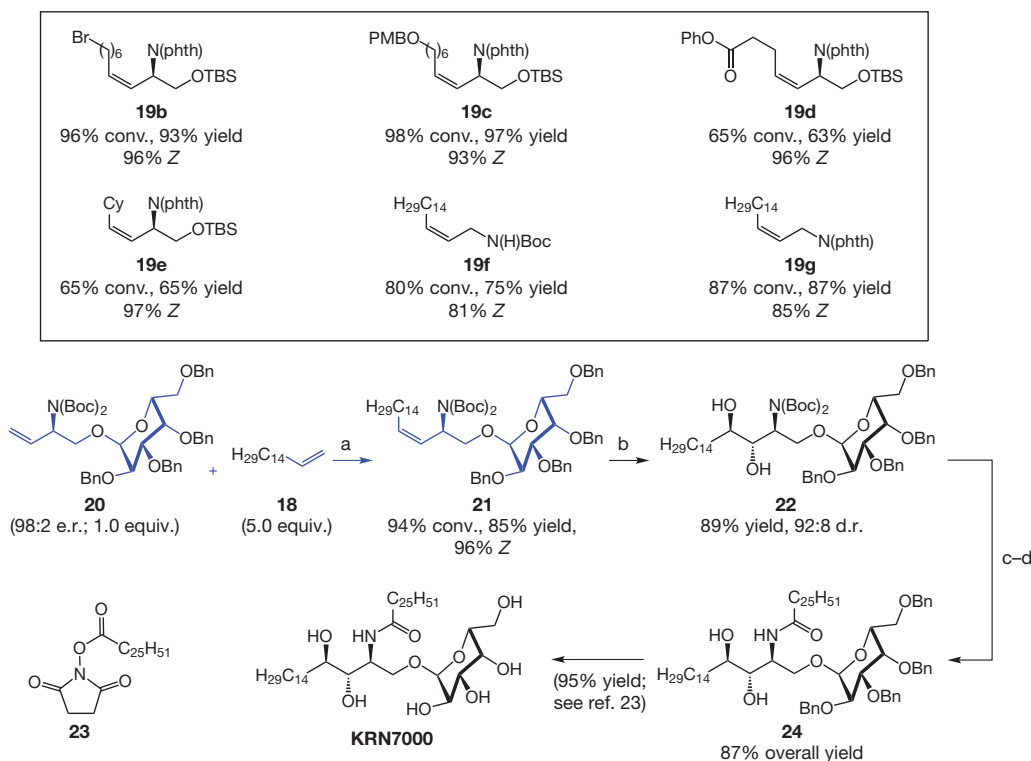
## Synthesis of natural product C18 (plasm)-16:0 (PC)

Next, we set out to demonstrate the utility of the catalytic CM process by a diastereo- and enantioselective synthesis of an anti-oxidant plasmalogen phospholipid, C18 (plasm)-16:0 (PC) (Fig. 2)<sup>16,17</sup>, the corresponding *E* isomer of which has been shown to be less active<sup>17</sup>. This initiative required addressing a challenge that is of general concern in catalytic CM: the inefficiency associated with the use of excess of one cross partner. The enol ether to be used (**11**) in the CM step is more valuable than the commercially available and inexpensive 1-octadecene (**12**), rendering utilization of excess amounts of the former unfavourable. Reducing the enol ether concentration diminishes efficiency and *Z*-selectivity, as detailed above and substantiated by the data in Table 2 (85% and 47% conversion with 5:1 and 1:1 **11:12**; entries 1 versus 2). Larger quantities of the less valuable **12** could improve yield and selectivity, as Mo-methylidene concentration is probably lowered through its reaction with excess alkene. However, increased amounts of an aliphatic alkene, unlike an enol ether, give rise to homocoupling and ethylene generation. Ethylene, in addition to being detrimental to the rate of CM (because it competes with the substrates for reaction with the available alkylidene), causes diminished stereoselectivity by increasing methylidene concentration, which promotes *Z* alkene isomerization (see above). We thus surmised that, if the negative effects of the generated ethylene were to be attenuated by performing the reaction



**Figure 2** | *Z*-selective CM reactions of enol ethers with terminal alkenes and application to stereoselective synthesis of C18 (plasm)-16:0 (PC). Various *Z* enol ethers are synthesized with 1.2–5.0 mol% of Mo complex **1a**, and typically require 2.0 equiv. (in the case of *p*-methoxyphenylvinyl ether) or 10.0 equiv. (with butylvinyl ether) of the terminal enol ether; excess butyl vinyl ether (**6**) is easily removed *in vacuo*. The desired *Z* alkenes are obtained in 51–77% yield and in 94% to >98% *Z* selectivity. Application to synthesis of C18 (plasm)-16:0 (PC) demonstrates the utility of the *Z*-selective Mo-catalysed CM, which is used in conjunction with a site- and enantioselective Cu-catalysed dihydroboration of the terminal alkyne in **14** (see Supplementary Information for details). All reactions shown were performed under N<sub>2</sub> atmosphere; catalysts were prepared

and used *in situ*. Conversions and *Z* selectivities were determined by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures; the variance of selectivity values is estimated to be <±2%. Yields of isolated products are shown (±5%). Reactions: for **8b–8e**, we used 2.5 mol% **1a** and 10 equiv. **6**; for **10a**, **10b**, we used 1.2 mol% **1a** and 2.0 equiv. **9**; for **10c**, **10d**, we used 5.0 mol% **1a** and 10 equiv. **9** (**10d**). See Supplementary Information for experimental details. Conditions for synthesis of **16**. Route a, step 1; 2.5 mol% **1a**, C<sub>6</sub>H<sub>6</sub>, 22 °C, 2.0 h, decalin, 1.0 torr; step 2; 5.0 equiv. (*n*-Bu)<sub>4</sub>NF, THF, 22 °C, 2 h. Route b; 2.5 mol% **15**, 2.5 mol% CuCl, 20 mol% NaOt-Bu, 2.1 equiv. *bis*(pinacolato)diboron, 3.0 equiv. MeOH, THF, 0 °C, 24 h; 30% H<sub>2</sub>O<sub>2</sub>, NaOH in aqueous THF, 1.0 h.



**Figure 3 | Z-selective CM reactions of allylic amides with terminal alkenes and application to stereoselective synthesis of KRN7000.** A range of Z-1,2-disubstituted allylic amides can be synthesized; in most cases, use of reduced pressure leads to substantially improved yield and stereoselectivity. Application to the stereoselective synthesis of KRN7000, involving catalytic diastereoselective dihydroxylation of the Z alkene obtained by Mo-catalysed CM, leads to an expeditious route for preparation of this biologically significant molecule (see Supplementary Information for details). All reactions shown were performed under N<sub>2</sub> atmosphere with 3.0 mol% **2**, 3.0 equiv. of the non-

N-containing alkenes (**19b**, **19c**) or 5.0 mol% **2** and 10.0 equiv. of cross partner, 7.0 torr, 5.0 hours, 22 °C; catalysts were prepared and used *in situ*. Conversions, Z selectivities, yields and Z:E ratios determined as in Fig. 1. For **19e**, reduced pressure was not used; reaction performed at 50 °C for 12 h. For **19f**, **19g**, reaction time was one hour. See Supplementary Information for experimental details. Conditions for synthesis of **24**: route a; 8.0 mol% **2** (*in situ*-generated), C<sub>6</sub>H<sub>6</sub>, 22 °C, 5.0 h, 1.0 torr. Route b; 5 mol% OsO<sub>4</sub>, 2.5 equiv. N-Me-morpholine oxide, CH<sub>2</sub>Cl<sub>2</sub>, 0 °C, 24 h. Route c; 10% trifluoroacetic acid, CH<sub>2</sub>Cl<sub>2</sub>, 22 °C, 30 min. Route d; 1.2 equiv. **23**, Et<sub>3</sub>N, THF, 50 °C, 12 h.

under vacuum, an efficient CM might be induced to proceed with only a relatively slight excess of the aliphatic alkene (**12**). Indeed, when catalytic CM is performed with an equal amount of **11** and **12** under 1.0 torr (entry 3, Table 2), efficiency (78% versus 47% conversion in entry 2) as well as stereoselectivity is substantially improved (97% versus 91.5% Z). With reduced pressure, 2 equiv. of **12** (versus **11**) and decalin as solvent (to prevent precipitation of the homocoupled by-product causing catalyst sequestration), 89% conversion is observed in two hours and Z-**13** is obtained with 97% selectivity (entry 4).

Removal of the silyl group delivers stereoisomerically pure Z-**14** in 85% overall yield (Fig. 2); the desired product cannot be accessed through catalytic hydrogenation of the corresponding alkyne (see also **10c**). Cu-catalysed site- and enantioselective dihydroboration<sup>18</sup> furnishes **16** (in 98:2 enantiomeric ratio, e.r.), which has been previously converted to

C18 (plasm)-16:0 (PC) in four steps and 86% overall yield<sup>19</sup>. Two additional points merit mention: (1) catalytic CM between **11** and **12** has been performed on the gram-scale with 1.0 mol% of *in situ*-generated **1a** and 2 equiv. of **12**, affording Z-**13** with >98% stereoselectivity and in 71% yield after purification (3 h, 1.0 torr, 79% conversion; see Supplementary Information for details). (2) The only previous synthesis of **16** involves nine transformations starting with (*S*)-isopropylidene glycerol (versus five reactions from (*i*-Pr)<sub>3</sub>Si-acetylene, Fig. 2) by a sequence that includes the use of highly toxic hexamethylphosphoramide and a catalytic hydrogenation with lead-containing salts<sup>19</sup>.

## Z-Selective cross-metathesis of allylic amides

Another class of reactions that we examined involves allylic amides as substrates. Such catalytic CM reactions are of considerable value, as a large number of biologically active molecules are nitrogen-containing, and 1,2-disubstituted alkenes bearing a C–N bond at the allylic position can be functionalized in a variety of ways. Furthermore, in contrast to transformations with enol ethers, CM with allylic amides poses the added complication that both substrates can undergo homocoupling. Preliminary investigations with enantiomerically pure allylic amide **17** (from commercially available alcohol) and 1-hexadecene **18** indicated that the optimal catalyst for this class of processes is derived from adamantylimido complex **2**, affording the desired Z alkene in 88% yield and with 97% stereoselectivity (entry 3, Table 3). Although arylimido derivatives **1a** and **1b** generate **19a** with similar selectivity (entries 1 and 2, Table 3), reactions are inefficient (26–44% versus 88% conversion), perhaps because an alkylidene derived from **2** is less congested and can more readily promote CM of the relatively hindered **17**. The higher efficiency of CM with **2**, in

**Table 2 | Effect of reduced pressure on efficiency and Z selectivity**

| Entry no. | 11:12 | Time (h) | Conv. (%) <sup>a</sup> | Solvent | Pressure | Z:E <sup>a</sup> |
|-----------|-------|----------|------------------------|---------|----------|------------------|
| 1         | 5:1   | 2        | 85                     | Benzene | Ambient  | >98:2            |
| 2         | 1:1   | 2        | 47                     | Benzene | Ambient  | 91.5:8.5         |
| 3         | 1:1   | 2        | 78                     | Benzene | 1.0 torr | 97:3             |
| 4         | 1:2   | 2        | 88                     | Decalin | 1.0 torr | 97:3             |

The reactions were carried out in purified benzene or decalin under an atmosphere of nitrogen gas (see Supplementary Information for details).

<sup>a</sup> Conversion, Z:E ratios and the amount of the homocoupled product were measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures; the variance of values is estimated to be  $\pm 2\%$ .



**Table 3** | Examination of various catalysts for CM with an allylic amide

| Entry no. | Complex   | Conv. (%) <sup>*</sup> | Yield (%) <sup>†</sup> | Z:E <sup>*</sup> |
|-----------|-----------|------------------------|------------------------|------------------|
| 1         | <b>1a</b> | 44                     | 35                     | 96:4             |
| 2         | <b>1b</b> | 26                     | 21                     | 97:3             |
| 3         | <b>2</b>  | 93                     | 88                     | 97:3             |
| 4         | <b>3</b>  | 9                      | 6                      | 21:79            |
| 5         | <b>4</b>  | 71                     | 68                     | 12:88            |
| 6         | <b>5</b>  | 73                     | 64                     | 11:89            |

The reactions were carried out in purified benzene under an atmosphere of nitrogen gas (see Supplementary Information for details). N(phth) = *N*-phthalimide.

<sup>\*</sup> Conversion and Z:E ratios were measured by analysis of 400 MHz <sup>1</sup>H NMR spectra of unpurified mixtures; the variance of values is estimated to be  $\pm 2\%$ .

<sup>†</sup> Yield of isolated product after purification; the variance of values is estimated to be  $\pm 5\%$ .

contrast to those involving enol ethers (Fig. 2), might be the result of CM with **17** being performed under vacuum, allowing minimal amounts of the relatively unstable methyldene to be formed. Chiral complex **3** is ineffective, and achiral Mo alkylidene **4** and Ru carbene **5** furnish the *E* isomer predominantly (79–89%). A weaker vacuum (7.0 torr versus 1.0 torr CM with enol ethers) is sufficient, indicating that such conditions can be applied to cases that involve relatively volatile substrates.

An assortment of allylic amides and terminal alkenes, including those that contain a halide (**19b**), a Lewis basic group (**19c**, **19d**) or a sterically demanding substituent (**19e**), can be used (Fig. 3). Stereoselective formation of **19f** and **19g** is noteworthy as the relatively less hindered unsaturated amides are more prone to homocoupling and the *Z* alkene products undergo equilibration to the *E* isomer readily, as manifested by the lower Z:E ratios. Although in certain cases 10 equiv. of a cross partner is used for maximum efficiency, lower amounts of alkene substrates lead to reasonably efficient processes. For example, with 3.0 mol% **2** and 3 equiv. of the aliphatic alkene (versus 5 mol% and 10 equiv.), **19g** is isolated in 62% yield and 90% *Z* selectivity (80% conversion, 5 min, 22 °C). It should be noted that in all the above transformations, use of catalysts bearing a racemic binol ligand furnishes similar levels of reactivity and stereoselectivity (see Supplementary Information).

## Synthesis of immunostimulant KRN7000

Stereoselective synthesis of anti-tumour agent KRN7000<sup>20,21</sup> underlines the utility of the method (Fig. 3). Catalytic CM of carbohydrate-containing allylic amide **20**, prepared in four steps from commercially available agents, affords **21** in 85% yield and with 96% *Z*-selectivity. Diastereoselective dihydroxylation (89% yield, 92:8 diastereomeric ratio (d.r.)) of the *Z* alkene delivers **22**; it should be noted that similar functionalization of the corresponding *E* alkene isomer would afford an undesired diol diastereomer<sup>22</sup>. Dihydroxylamide **24** is secured in two steps and the target is obtained after carbohydrate deprotection<sup>23</sup>. *Z*-Selective CM thus provides access to a route that is significantly more concise than the 14-step sequence (compared to nine steps in Fig. 3) reported thus far as the shortest synthesis of KRN7000<sup>24</sup>. It is noteworthy that the convergent nature of a synthesis approach involving catalytic CM, such as the two examples provided here, can easily translate to preparation of a variety of related analogues; for example, in connection with preparation of **21** (Fig. 3), a wide range of other terminal alkenes may be used.

## The balance between conversion and *Z* selectivity

The relationship between efficiency and stereoselectivity is critical and merits a brief discussion. The conversion values, at times less than complete, represent a balance struck between achieving the highest yield and maximal *Z* selectivity with minimal substrate equivalents

and adventitious homocoupling. Transformations performed under ambient conditions (no vacuum) may not proceed beyond 80% conversion, probably because the ethylene by-product competes with the remaining cross partner molecules. High ethylene concentration might also diminish CM rate through formation of relatively stable unsubstituted metallacyclobutanes<sup>7</sup>. As mentioned before, methyldene complexes can engender reduction of *Z* selectivity; time-dependent studies indicate that stereoselectivities suffer with prolonged reaction times. With non-volatile substrates, if reactions are carried out under vacuum, complete consumption of the limiting alkene is observed only when excess amounts (~10 equiv.) of one cross partner are present. Under such regimes, however, difficulties associated with removal of the excess substrate and the homocoupled product might arise, rendering the use of lesser alkene amounts preferable. With lower substrate ratios, >98% consumption of the limiting substrate is difficult to achieve, as terminal alkene concentration is diminished as a result of partial homocoupling.

## Conclusions and discussions

In addition to catalytic olefin CM, Wittig reactions<sup>25</sup>, catalytic alkyne hydrogenation and cross-coupling<sup>1,26</sup> are notable approaches for synthesis of *Z*-disubstituted alkenes. The above four types of transformations are distinct—each delivers the desired product through a different bond disconnection. Similarly to CM, in cross-coupling alkenes serve as starting materials; in contrast to CM, however, it is through the synthesis of the substrate (for example, a *Z* vinyl halide)—and not in the cross-coupling step—that the stereochemical identity of the product is determined. Wittig-type processes are typically not catalytic and involve reaction of aldehydes (versus the more stable alkenes) and triphenylphosphonium ylides. Catalytic alkyne hydrogenation requires substrates derived from functionalization of a terminal alkyne; currently, relative to alkenes, methods for preparation of alkynes are less common and related synthesis routes are often lengthier. Moreover, partial hydrogenation of alkynes involves metal catalysts that contain poisonous lead salts and must be controlled to avoid over-reduction and generation of alkane by-products that can be difficult to separate from the desired *Z* alkene. Catalytic CM thus offers a desirable alternative to synthesis of *Z* alkenes, particularly as it requires as starting material a functional group that is stable, easily accessible and distinct from the other commonly used protocols mentioned above.

The strategies outlined here—including the use of reduced pressure to enhance stereoselectivity in catalytic CM, and the *Z*-selective Mo-catalysed transformations—offer a unique solution to a long-standing problem in organic chemistry<sup>27</sup>. Our findings offer additional evidence regarding the unique ability of stereogenic-at-Mo mono-aryloxyppyrolides to effect olefin metathesis reactions that extend beyond enantioselective processes<sup>28</sup>, with efficiency and selectivity levels that are not achievable with other catalyst classes. The catalytic processes described here are expected to affect significantly activities that require the stereoselective synthesis of organic molecules<sup>29,30</sup>.

## METHODS SUMMARY

**General procedure for catalytic *Z*-selective cross-metathesis.** In an N<sub>2</sub>-filled dry box, an oven-dried (135 °C) 20-ml vial equipped with a magnetic stir bar was charged with vinyl ether **11** (1.00 g, 4.19 mmol) and 1.0 mol% of *in situ*-generated complex **1a** (419 μl, 0.100 M, 41.9 μmol; final substrate concentration = 1.70 M). A separate 2.0-ml vial was charged with 1-octadecene (**12**, 2.12 g, 8.39 mmol) and decalin (2.10 ml). The resulting solution was transferred to the mixture of **11** and **1a** by syringe; a septum, fitted with an outlet needle, was attached to the vial and an adapter was attached to the top of the septum and vacuum (~1.0 torr) applied. The resulting solution was stirred for 3 h. The vessel was removed from the dry box and the reaction was quenched by the addition of wet Et<sub>2</sub>O (~1.0 ml). The unpurified product is >98% *Z* (as determined by 400 MHz <sup>1</sup>H NMR analysis). The residue was dissolved in Et<sub>2</sub>O and passed through a 2.5-cm plug of neutral alumina to remove inorganic salts, and the solution was concentrated. In a 25-ml round-bottom flask equipped with a stir bar, the resulting residue was treated

with (*n*-Bu)<sub>4</sub>NF (1.0 M in THF, 21.0 ml, 21.0 mmol), and stirred for 2 h. The mixture was diluted with Et<sub>2</sub>O (200 ml), passed through a 5-cm plug of neutral alumina, and concentrated. The resulting white solid was purified by chromatography on neutral alumina (100% hexanes) to afford **15** as a white solid (m.p. 30–31 °C, 0.914 g, 2.98 mmol, 71.0% yield; >98% Z isomer).

Received 25 January; accepted 22 February 2011.

- Negishi, E.-i. *et al.* Recent advances in efficient and selective synthesis of di-, tri-, and tetrasubstituted alkenes via Pd-catalyzed alkenylation–carbonyl olefination strategy. *Acc. Chem. Res.* **41**, 1474–1485 (2008).
- Chatterjee, A. K. in *Handbook of Metathesis* Vol. 1 (ed. Grubbs, R. H.) 246–295 (Wiley-VCH, 2003).
- Crowe, W. E. & Goldberg, D. R. Acrylonitrile cross-metathesis: coaxing olefin metathesis reactivity from a reluctant substrate. *J. Am. Chem. Soc.* **117**, 5162–5163 (1995).
- Hansen, E. C. & Lee, D. Efficient and Z-selective cross-metathesis of conjugated enynes. *Org. Lett.* **6**, 2035–2038 (2004).
- Malcolmson, S. J., Meek, S. J., Sattely, E. S., Schrock, R. R. & Hoveyda, A. H. Highly efficient molybdenum-based catalysts for alkene metathesis. *Nature* **456**, 933–937 (2008).
- Lee, Y.-J., Schrock, R. R. & Hoveyda, A. H. Endo-selective enyne ring-closing metathesis promoted by stereogenic-at-Mo monoalkoxide and monoaryloxide complexes. Efficient synthesis of cyclic dienes not accessible through reactions with Ru carbenes. *J. Am. Chem. Soc.* **131**, 10652–10661 (2009).
- Solans-Monfort, X., Clot, E., Copéret, C. & Eisenstein, O. d<sup>0</sup>-Re-based olefin metathesis catalysts, Re(=CR)(=CHR)(X)(Y): the key role of X and Y ligands for efficient active sites. *J. Am. Chem. Soc.* **127**, 14015–14025 (2005).
- Meek, S. J., Malcolmson, S. J., Li, B., Schrock, R. R. & Hoveyda, A. H. The significance of degenerate processes to enantioselective olefin metathesis reactions promoted by stereogenic-at-Mo complexes. *J. Am. Chem. Soc.* **131**, 16407–16409 (2009).
- Alexander, J. B., La, D. S., Cefalo, D. R., Hoveyda, A. H. & Schrock, R. R. Catalytic enantioselective ring-closing metathesis by a chiral biphen-Mo complex. *J. Am. Chem. Soc.* **120**, 4041–4042 (1998).
- Schrock, R. R. & Hoveyda, A. H. Molybdenum and tungsten imido alkylidene complexes as efficient olefin metathesis catalysts. *Angew. Chem. Int. Edn* **42**, 4592–4633 (2003).
- Garber, S. B., Kingsbury, J. S., Gray, B. L. & Hoveyda, A. H. Efficient and recyclable monomeric and dendritic Ru-based metathesis catalysts. *J. Am. Chem. Soc.* **122**, 8168–8179 (2000).
- Ibrahim, I., Yu, M., Schrock, R. R. & Hoveyda, A. H. Highly Z- and enantioselective ring-opening/cross-metathesis reactions catalyzed by stereogenic-at-Mo adamantylimido complexes. *J. Am. Chem. Soc.* **131**, 3844–3845 (2009).
- Jiang, A. J., Zhao, Y., Schrock, R. R. & Hoveyda, A. H. Highly Z-selective metathesis homocoupling of terminal olefins. *J. Am. Chem. Soc.* **131**, 16630–16631 (2009).
- Schrader, T. O. & Snapper, M. L. in *Handbook of Metathesis* (ed. Grubbs, R. H.) 205–237 (Wiley-VCH, 2003).
- La, D. S., Sattely, E. S., Ford, J. G., Schrock, R. R. & Hoveyda, A. H. Catalytic asymmetric ring-opening metathesis/cross metathesis (AROM/CM) reactions. Mechanism and application to enantioselective synthesis of functionalized cyclopentanes. *J. Am. Chem. Soc.* **123**, 7767–7778 (2001).
- Nagan, N. & Zoeller, R. A. Plasmalogens: biosynthesis and functions. *Prog. Lipid Res.* **40**, 199–229 (2001).
- Lankalapalli, R. S. *et al.* Synthesis and antioxidant properties of an unnatural plasmalogen analogue bearing a trans O-vinyl ether linkage. *Org. Lett.* **11**, 2784–2787 (2009).
- Lee, Y., Jang, H. & Hoveyda, A. H. Vicinal diboronates in high enantiomeric purity through tandem site-selective NHC–Cu-catalyzed boron–copper additions to terminal alkynes. *J. Am. Chem. Soc.* **131**, 18234–18235 (2009).
- Qin, D., Byun, H.-S. & Bittman, R. Synthesis of plasmalogen via 2,3-bis-O-(4'-methoxybenzyl)-sn-glycerol. *J. Am. Chem. Soc.* **121**, 662–668 (1999).
- Borg, N. A. *et al.* CD1d-lipid-antigen recognition by the semi-invariant NKT T-cell receptor. *Nature* **448**, 44–49 (2007).
- Trappeniens, M. *et al.* Synthesis and in vitro evaluation of α-GalCer epimers. *ChemMedChem* **3**, 1061–1070 (2008).
- Llaveria, J., Díaz, Y., Matheu, I. & Castillón, S. An efficient and general enantioselective synthesis of sphingosine, phytosphingosine, and 4-substituted derivatives. *Org. Lett.* **11**, 205–208 (2009).
- Kim, S., Song, S., Lee, T., Jung, S. & Kim, D. Practical synthesis of KRN7000 from phytosphingosine. *Synthesis* 847–850 (2004).
- Michieletti, M. *et al.* Synthesis of α-galactosyl ceramide (KRN7000) and analogues thereof via a common precursor and their preliminary biological assessment. *J. Org. Chem.* **73**, 9192–9195 (2008).
- Nicolaou, K. C., Härter, M. W., Gunzner, J. L. & Nadin, A. The Wittig and related reactions in natural product synthesis. *Justus Liebigs Ann. Chem.* 1283–1301 (1997).
- Miyaura, N. & Suzuki, A. Palladium-catalyzed cross-coupling reactions of organoboron compounds. *Chem. Rev.* **95**, 2457–2483 (1995).
- Hoveyda, A. H. & Zhugralin, A. R. The remarkable metal-catalysed olefin metathesis reaction. *Nature* **450**, 243–251 (2007).
- Hoveyda, A. H., Malcolmson, S. J., Meek, S. J. & Zhugralin, A. R. Catalytic enantioselective olefin metathesis in natural product synthesis. Chiral metal-based complexes that deliver high enantioselectivity and more. *Angew. Chem. Int. Edn* **49**, 34–44 (2010).
- Nicolaou, K. C., Bulger, P. G. & Sarlah, D. Metathesis reactions in total synthesis. *Angew. Chem. Int. Edn* **44**, 4490–4527 (2005).
- Prunet, J. & Grimaud, L. in *Metathesis in Natural Product Synthesis* (eds Cossy, J., Arseniyadis, S. & Meyer, C.) 287–312 (Wiley-VCH, 2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This research was supported by the US National Institutes of Health, Institute of General Medical Sciences (GM-59426 to A.H.H. and R.R.S.) and the National Science Foundation (CHE-0715138 to A.H.H.). R.V.O. and J.L. were LaMattina and Spanish government Visiting Scholar Fellows, respectively. We thank S. Castillón, S. J. Malcolmson and M. Yu for discussions. Mass spectrometry facilities at Boston College are supported by the US National Science Foundation (DBI-0619576).

**Author Contributions** S.J.M., R.V.O. and J.L. were involved in the discovery, design and development of the new Z-selective cross-metathesis strategies and applications to the natural product syntheses. A.H.H. and R.R.S. conceived the research programme. A.H.H. directed the investigations and composed the manuscript, with revisions provided by S.J.M. and R.V.O.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests; details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.H.H. ([amir.hoveyda@bc.edu](mailto:amir.hoveyda@bc.edu)).



# Initial genome sequencing and analysis of multiple myeloma

Michael A. Chapman<sup>1†</sup>, Michael S. Lawrence<sup>1</sup>, Jonathan J. Keats<sup>2,3</sup>, Kristian Cibulskis<sup>1</sup>, Carrie Sougnez<sup>1</sup>, Anna C. Schinzel<sup>4</sup>, Christina L. Harview<sup>1</sup>, Jean-Philippe Brunet<sup>1</sup>, Gregory J. Ahmann<sup>2,3</sup>, Mazhar Adli<sup>1,5</sup>, Kenneth C. Anderson<sup>3,4</sup>, Kristin G. Ardlie<sup>1</sup>, Daniel Auclair<sup>3,6</sup>, Angela Baker<sup>7</sup>, P. Leif Bergsagel<sup>2,3</sup>, Bradley E. Bernstein<sup>1,5,8,9</sup>, Yotam Drier<sup>1,10</sup>, Rafael Fonseca<sup>2,3</sup>, Stacey B. Gabriel<sup>1</sup>, Craig C. Hofmeister<sup>3,11</sup>, Sundar Jagannath<sup>3,12</sup>, Andrzej J. Jakubowiak<sup>3,13</sup>, Amrita Krishnan<sup>3,14</sup>, Joan Levy<sup>3,6</sup>, Ted Liefeld<sup>1</sup>, Sagar Lonial<sup>3,15</sup>, Scott Mahan<sup>1</sup>, Bunmi Mfuko<sup>3,6</sup>, Stefano Monti<sup>1</sup>, Louise M. Perkins<sup>3,6</sup>, Robb Onofrio<sup>1</sup>, Trevor J. Pugh<sup>1</sup>, S. Vincent Rajkumar<sup>3,16</sup>, Alex H. Ramos<sup>1</sup>, David S. Siegel<sup>3,17</sup>, Andrey Sivachenko<sup>1</sup>, A. Keith Stewart<sup>2,3</sup>, Suzanne Trudel<sup>3,18</sup>, Ravi Vij<sup>3,19</sup>, Douglas Voet<sup>1</sup>, Wendy Winckler<sup>1</sup>, Todd Zimmerman<sup>3,20</sup>, John Carpten<sup>7</sup>, Jeff Trent<sup>7</sup>, William C. Hahn<sup>1,4,8</sup>, Levi A. Garraway<sup>1,4</sup>, Matthew Meyerson<sup>1,4,8</sup>, Eric S. Lander<sup>1,8,21</sup>, Gad Getz<sup>1</sup> & Todd R. Golub<sup>1,4,8,9</sup>

**Multiple myeloma is an incurable malignancy of plasma cells, and its pathogenesis is poorly understood. Here we report the massively parallel sequencing of 38 tumour genomes and their comparison to matched normal DNAs. Several new and unexpected oncogenic mechanisms were suggested by the pattern of somatic mutation across the data set. These include the mutation of genes involved in protein translation (seen in nearly half of the patients), genes involved in histone methylation, and genes involved in blood coagulation. In addition, a broader than anticipated role of NF- $\kappa$ B signalling was indicated by mutations in 11 members of the NF- $\kappa$ B pathway. Of potential immediate clinical relevance, activating mutations of the kinase BRAF were observed in 4% of patients, suggesting the evaluation of BRAF inhibitors in multiple myeloma clinical trials. These results indicate that cancer genome sequencing of large collections of samples will yield new insights into cancer not anticipated by existing knowledge.**

Multiple myeloma is an incurable malignancy of mature B-lymphoid cells, and its pathogenesis is only partially understood. About 40% of cases harbour chromosome translocations resulting in overexpression of genes (including *CCND1*, *CCND3*, *MAF*, *MAFB*, *WHSC1* (also called *MMSET*) and *FGFR3*) via their juxtaposition to the immunoglobulin heavy chain (IgH) locus<sup>1</sup>. Other cases exhibit hyperdiploidy. However, these abnormalities are probably insufficient for malignant transformation because they are also observed in the pre-malignant syndrome known as monoclonal gammopathy of uncertain significance. Malignant progression events include activation of *MYC*, *FGFR3*, *KRAS* and *NRAS* and activation of the NF- $\kappa$ B pathway<sup>1–3</sup>. More recently, loss-of-function mutations in the histone demethylase *UTX* (also called *KDM6A*) have also been reported<sup>4</sup>.

A powerful way to understand the molecular basis of cancer is to sequence either the entire genome or the protein-coding exome, comparing tumour to normal from the same patient to identify the acquired somatic mutations. Recent reports have described the sequencing of whole genomes from a single patient<sup>5–9</sup>. Although informative, we hypothesized that a larger number of cases would permit the identification of biologically relevant patterns that would not otherwise be evident.

## Landscape of multiple myeloma mutations

We studied 38 multiple myeloma patients (Supplementary Table 1), performing whole-genome sequencing (WGS) for 23 patients and

whole-exome sequencing (WES; assessing 164,687 exons) for 16 patients, with one patient analysed by both approaches (Supplementary Information). WES is a cost-effective strategy to identify protein-coding mutations, but cannot detect non-coding mutations and rearrangements. We identified tumour-specific mutations by comparing each tumour to its corresponding normal, using a series of algorithms designed to detect point mutations, small insertions/deletions (indels) and other rearrangements (Supplementary Fig. 1). On the basis of WGS, the frequency of tumour-specific point mutations was 2.9 per million bases, corresponding to approximately 7,450 point mutations per sample across the genome, including an average of 35 amino-acid-changing point mutations plus 21 chromosomal rearrangements disrupting protein-coding regions (Supplementary Tables 2 and 3). The mutation-calling algorithm was found to be highly accurate, with a true positive rate of 95% for point mutations (Supplementary Text, Supplementary Tables 4 and 5, and Supplementary Fig. 2).

The mutation rate across the genome varied greatly depending on base composition, with mutations at CpG dinucleotides occurring fourfold more commonly than mutations at A or T bases (Supplementary Fig. 3a). In addition, even after correction for base composition, the mutation frequency in coding regions was lower than that observed in intronic and intergenic regions ( $P < 1 \times 10^{-16}$ ; Supplementary Fig. 3b), potentially owing to negative selective pressure against mutations disrupting coding sequences. There is also a

<sup>1</sup>The Eli and Edythe L. Broad Institute, 7 Cambridge Center, Cambridge, Massachusetts 02412, USA. <sup>2</sup>Mayo Clinic Arizona, 13400 East Shea Boulevard, Scottsdale, Arizona 85259, USA. <sup>3</sup>The Multiple Myeloma Research Consortium, 383 Main Avenue, 5th Floor, Norwalk, Connecticut 06851, USA. <sup>4</sup>Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA. <sup>5</sup>Massachusetts General Hospital, 55 Fruit Street, Boston, Massachusetts 02114, USA. <sup>6</sup>Multiple Myeloma Research Foundation, 383 Main Avenue, 5th Floor, Norwalk, Connecticut 06851, USA. <sup>7</sup>The Translational Genomics Research Institute, 445 North Fifth Street, Phoenix, Arizona 85004, USA. <sup>8</sup>Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>9</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA. <sup>10</sup>Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. <sup>11</sup>The Ohio State University Medical Center, 320 West 10th Avenue, Columbus, Ohio 43210, USA. <sup>12</sup>St Vincent's Comprehensive Cancer Center, 325 West 15th Street, New York, New York 11001, USA. <sup>13</sup>University of Michigan Comprehensive Cancer Center, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA. <sup>14</sup>City of Hope Comprehensive Cancer Center, 1500 East Duarte Road, Duarte, California 91010, USA. <sup>15</sup>Winship Cancer Institute, Emory University, 1365-C Clifton Road, NE Atlanta, Georgia 30322, USA. <sup>16</sup>Mayo Clinic Rochester, 200 1st Street SW Rochester, Minnesota 55905, USA. <sup>17</sup>John Theurer Cancer Center at Hackensack University Medical Center, 92 Second Street, Hackensack, New Jersey 07601, USA. <sup>18</sup>Princess Margaret Hospital, 610 University Avenue, Toronto, Ontario M5G 2M9, Canada. <sup>19</sup>Washington University School of Medicine, Washington University in St Louis, 660 South Euclid Avenue, St Louis, Missouri 63110, USA. <sup>20</sup>University of Chicago Medical Center, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA. <sup>21</sup>Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. †Present address: Department of Haematology, Addenbrookes Hospital, Hills Road, Cambridge CB2 0QQ, UK.

lower mutation rate in intronic regions compared to intergenic regions ( $P < 1 \times 10^{-16}$ ), which may reflect transcription-coupled repair, as previously suggested<sup>10,11</sup>. Consistent with this explanation, we observed a lower mutation rate in introns of genes expressed in multiple myeloma compared to those not expressed (Fig. 1a).

### Frequently mutated genes

We next focused on the distribution of somatic, non-silent protein-coding mutations. We estimated statistical significance by comparison to the background distribution of mutations (Supplementary Information). Ten genes showed statistically significant rates of protein-altering mutations ('significantly mutated genes') at a false discovery rate (FDR) of  $\leq 0.10$  (Table 1). To investigate their functional importance, we compared their predicted consequence (on the basis of evolutionary conservation and nature of the amino acid

change) to the distribution of all coding mutations. This analysis showed a dramatic skewing of functional importance (FI) scores<sup>12</sup> for the ten significantly mutated genes ( $P = 7.6 \times 10^{-14}$ ; Fig. 1b), supporting their biological relevance. Even after RAS and p53 mutations are excluded from the analysis, the skewing remained significant ( $P < 0.01$ ).

We also examined the non-synonymous/synonymous (NS/S) mutation rate for the significantly mutated genes. The expected NS/S ratio was  $2.82 \pm 0.15$ , whereas the observed ratio was 39:0 for the significant genes ( $P < 0.0001$ ), further strengthening the case that these genes are probably drivers of the pathogenesis of multiple myeloma, and are unlikely to simply be passenger mutations.

The significantly mutated genes include three previously reported to have point mutations in multiple myeloma: *KRAS* and *NRAS* (10 and 9 cases, respectively (50%,  $P < 1 \times 10^{-11}$ ,  $q < 1 \times 10^{-6}$ ), and *TP53* (3 cases (8%),  $P = 5.1 \times 10^{-6}$ ,  $q = 0.019$ ). Interestingly, we identified two point mutations (5%,  $P = 0.000027$ ,  $q = 0.086$ ) in *CCND1* (cyclin D1), which has long been recognized as a target of chromosomal translocation in multiple myeloma, but for which point mutations have not been observed previously in cancer.

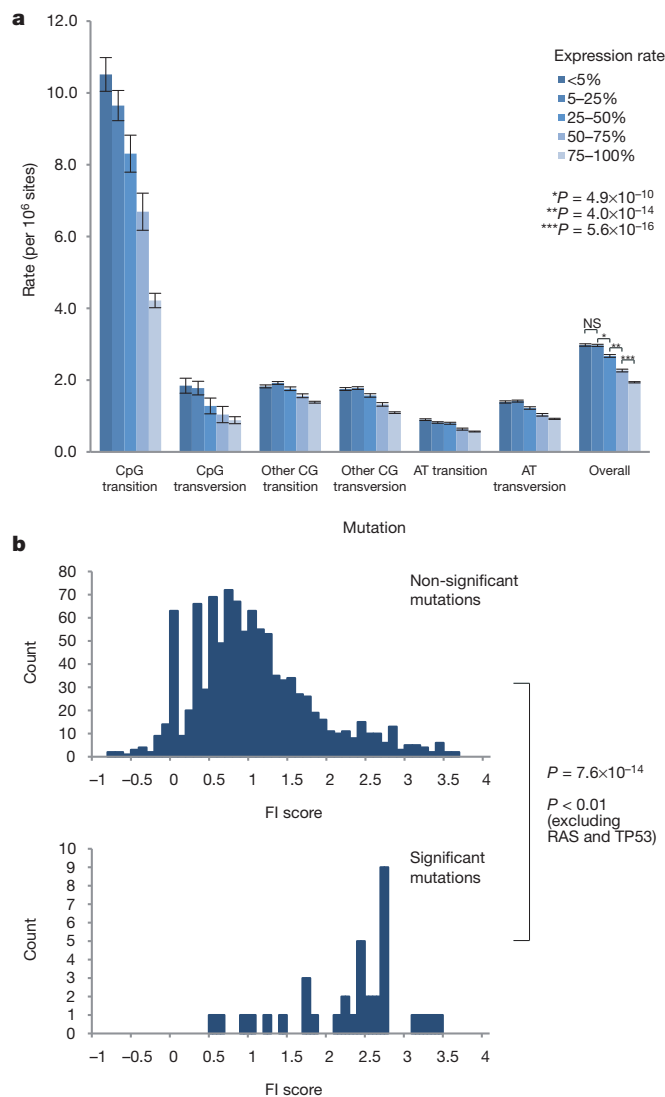
The remaining six genes have not previously been known to be involved in cancer, and indicate new aspects of the pathogenesis of multiple myeloma.

### RNA processing and protein homeostasis mutations

A striking finding of this study was the discovery of frequent mutations in genes involved in RNA processing, protein translation and the unfolded protein response. Such mutations were observed in nearly half of the patients.

The *DIS3* (also called *RRP44*) gene harboured mutations in 4 out of 38 patients (11%,  $P = 2.4 \times 10^{-6}$ ,  $q = 0.011$ ). *DIS3* encodes a highly conserved RNA exonuclease which serves as the catalytic component of the exosome complex involved in regulating the processing and abundance of all RNA species<sup>13,14</sup>. The four observed mutations occur at highly conserved regions (Fig. 2a) and cluster within the RNB domain facing the enzyme's catalytic pocket (Fig. 2b). Two lines of evidence indicate that the *DIS3* mutations result in loss of function. First, three of the four tumours with mutations exhibited loss of heterozygosity via deletion of the remaining *DIS3* allele. Second, two of the mutations have been functionally characterized in yeast and bacteria, where they result in loss of enzymatic activity leading to the accumulation of their RNA targets<sup>15,16</sup>. Given that a key role of the exosome is the regulation of the available pool of mRNAs available for translation<sup>17</sup>, these results indicate that *DIS3* mutations may dysregulate protein translation as an oncogenic mechanism in multiple myeloma.

Further support for a role of translational control in the pathogenesis of multiple myeloma comes from the observation of mutations in the *FAM46C* gene in 5 out of 38 (13%) patients ( $P = 1.8 \times 10^{-10}$ ,  $q = 1 \times 10^{-6}$ ). There is no published functional annotation of *FAM46C*, and its sequence lacks obvious homology to known proteins. To gain insight into its cellular role, we examined its pattern of gene expression across 414 multiple myeloma samples and compared it to the expression of 395 gene sets curated in the Molecular Signatures Database (MSigDB), using the GSEA algorithm<sup>18–20</sup>. The expression of *FAM46C* was highly correlated ( $q = 0.034$  after multiple hypothesis correction; Fig. 2c) to the expression of the set of ribosomal proteins that are known to be tightly co-regulated<sup>21</sup>. Strong correlation with eukaryotic initiation and elongation factors involved in protein translation was similarly observed. Although the precise function of *FAM46C* remains unknown, this striking correlation provides strong evidence that *FAM46C* is functionally related in some way to the regulation of translation. Consistent with this observation, *FAM46C* was recently shown to function as an mRNA stability factor (M. Fleming, manuscript submitted).



**Figure 1 | Evidence for transcription-coupled repair and functional importance of statistically significant mutations.** **a**, Intronic mutation rates subdivided by gene expression rates in multiple myeloma. Rates of gene expression were estimated by proportion of Affymetrix Percent (P) calls in 304 primary multiple myeloma samples. Error bars indicate  $\pm 1$  standard deviation. NS, not significant. **b**, Functional importance (FI) scores were generated for all point mutations and divided into distributions for nonsignificant mutations (top histogram;  $n = 1,019$ ) and significant mutations (bottom;  $n = 36$ ). Comparison of distributions is via the Kolmogorov–Smirnov statistic.



**Table 1 | Statistically significant protein-coding mutations in multiple myeloma**

| Gene           | N      | n  | Untreated n | CpG transition | Other C:G transition | C:G transversion | A:T mutation | Indel/ null | P-value                | q-value               |
|----------------|--------|----|-------------|----------------|----------------------|------------------|--------------|-------------|------------------------|-----------------------|
| <i>NRAS</i>    | 20,711 | 9  | 3           | 0              | 0                    | 3                | 6            | 0           | $<1.0 \times 10^{-11}$ | $<1.0 \times 10^{-6}$ |
| <i>KRAS</i>    | 25,728 | 10 | 6           | 0              | 5                    | 1                | 4            | 0           | $<1.0 \times 10^{-11}$ | $<1.0 \times 10^{-6}$ |
| <i>FAM46C</i>  | 39,661 | 5  | 3           | 0              | 0                    | 2                | 1            | 2           | $1.8 \times 10^{-10}$  | $1.0 \times 10^{-6}$  |
| <i>DIS3</i>    | 89,758 | 4  | 1           | 0              | 1                    | 1                | 2            | 0           | $2.4 \times 10^{-6}$   | 0.011                 |
| <i>TP53</i>    | 32,585 | 3  | 1           | 0              | 0                    | 1                | 1            | 1           | $5.1 \times 10^{-6}$   | 0.019                 |
| <i>CCND1</i>   | 12,899 | 2  | 1           | 0              | 0                    | 0                | 2            | 0           | 0.00027                | 0.086                 |
| <i>PNRC1</i>   | 19,621 | 2  | 2           | 0              | 1                    | 0                | 0            | 1           | 0.00039                | 0.094                 |
| <i>ALOX12B</i> | 40,369 | 3  | 0           | 1              | 0                    | 1                | 1            | 0           | 0.00042                | 0.094                 |
| <i>HLA-A</i>   | 18,635 | 2  | 0           | 0              | 0                    | 0                | 2            | 0           | 0.00045                | 0.094                 |
| <i>MAGED1</i>  | 53,950 | 2  | 1           | 0              | 0                    | 0                | 0            | 2           | 0.00053                | 0.10                  |

Territory (N) refers to total covered territory in base pairs across 38 sequenced samples. Total numbers of mutations (n) and numbers of mutations occurring in therapy-naïve disease (Untreated n) are shown for each gene.

Notably, although not statistically significant on their own, we found mutations in five other genes related to protein translation, stability and the unfolded protein responses (Supplementary Table 6), further supporting a role of translational control in multiple myeloma. Of particular interest, two patients had mutations in the unfolded protein response gene *XBPI*. Overexpression of a particular splice form of *XBPI* has been shown to cause a multiple-myeloma-like syndrome in mice, although no role of *XBPI* in the pathogenesis of human multiple myeloma has been described<sup>22</sup>.

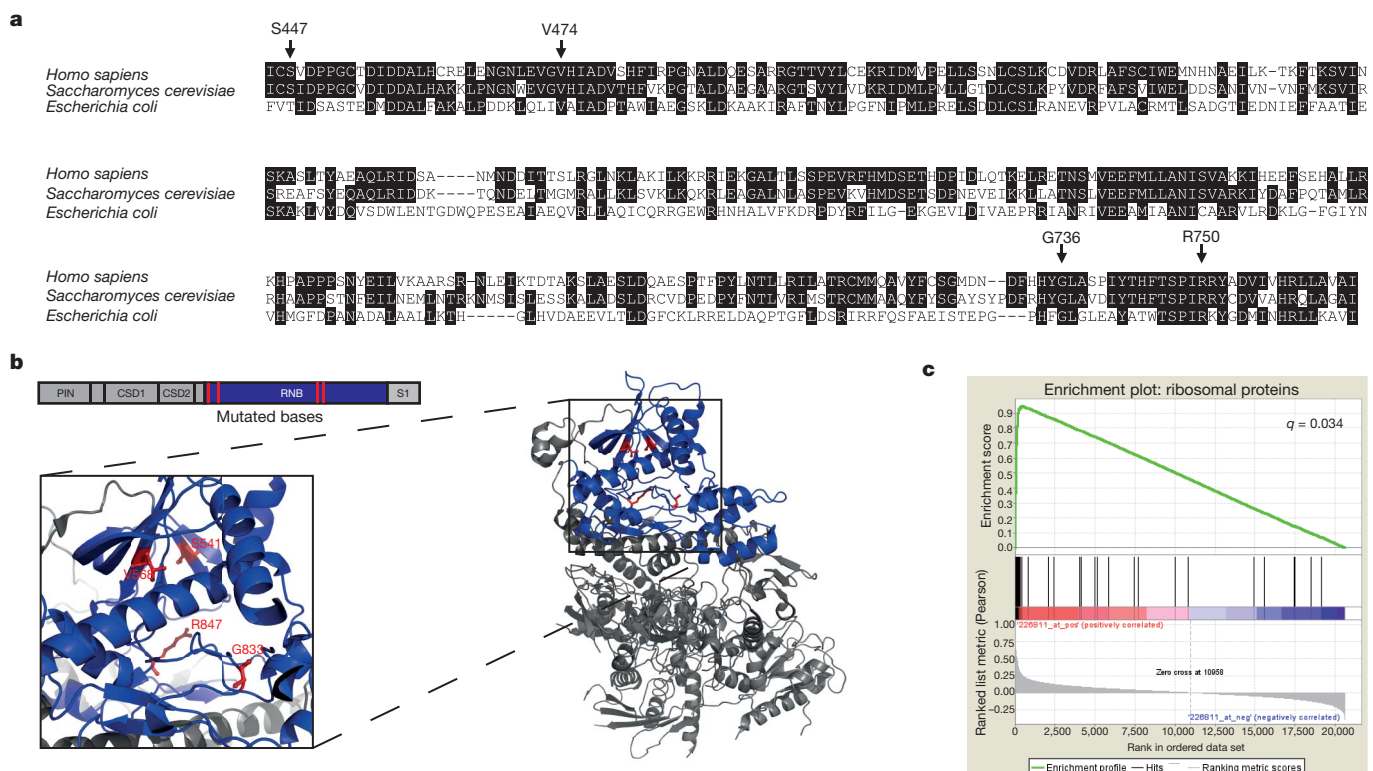
Of related interest, mutations of the *LRRK2* gene were observed in 3 out of 38 patients (8%; Supplementary Table 6). *LRRK2* encodes a serine-threonine kinase that phosphorylates translation initiation factor 4E-binding protein (4EBP). *LRRK2* is best known for its role in the predisposition to Parkinson's disease<sup>23,24</sup>. Parkinson's disease and other neurodegenerative diseases such as Huntington's disease are characterized in part by aberrant unfolded protein responses<sup>25</sup>. Protein homeostasis may be particularly important in multiple myeloma because of the enormous rate of production of immunoglobulins by

multiple myeloma cells<sup>26–28</sup>. The finding is also of clinical significance because of the success of the drug bortezomib (Velcade), which inhibits the proteasome and which shows remarkable activity in multiple myeloma compared to other tumour types<sup>29</sup>.

Together, these results indicate that mutations affecting protein translation and homeostasis are extremely common in multiple myeloma (at least 16 out of 38 patients; 42%), thereby indicating that additional therapeutic approaches that target these mechanisms may be worth exploring.

### Identical mutations suggest gain-of-function oncogenes

Another way to recognize biologically significant mutations is to search for recurrence of identical mutations indicative of gain-of-function alterations in oncogenes. Two patients had an identical mutation (K123R) in the DNA-binding domain of the interferon regulatory factor IRF4. Interestingly, a recent RNA interference screen in multiple myeloma showed that IRF4 was required for multiple myeloma survival, consistent with its role as a putative oncogene<sup>30</sup>. Genotyping for



**Figure 2 | Mutations likely to affect protein translation and/or homeostasis in multiple myeloma.** **a**, Alignment of human, yeast and bacterial RNB domain of DIS3. Positions of observed mutations are indicated with respect to the human sequence. Yeast equivalents are, respectively, S541, V568, G833 and

R847. **b**, One-dimensional and three-dimensional structures of yeast DIS3, with the RNB domain coloured in blue and mutations coloured in red. **c**, GSEA plot showing enrichment of ribosomal protein gene set among genes correlated with *FAM46C* expression in 414 multiple myeloma samples.

this mutation in 161 additional multiple myeloma samples identified two more patients with this mutation. IRF4 is a transcriptional regulator of PRDM1 (also called BLIMP1), and two of 38 sequenced patients also exhibited *PRDM1* mutations. PRDM1 is a transcription factor involved in plasma cell differentiation, loss-of-function mutations of which occur in diffuse large B-cell lymphoma<sup>31–35</sup>.

### Clinically actionable mutations in BRAF

Some mutations deserve attention because of their clinical relevance. One of the thirty-eight patients harboured a BRAF kinase mutation (G469A). Although BRAF G469A has not previously been observed in multiple myeloma, this precise mutation is known to be activating and oncogenic<sup>36</sup>. We genotyped an additional 161 multiple myeloma patients for the 12 most common *BRAF* mutations and found mutations in 7 patients (4%). Three of these were K601N and four were V600E (the most common BRAF mutation in melanoma<sup>37</sup>). Our finding of common *BRAF* mutations in multiple myeloma has important clinical implications because such patients may benefit from treatment with BRAF inhibitors, some of which show marked clinical activity<sup>38</sup>. Our results also support the observation that inhibitors acting downstream of BRAF (for example, on MEK) may have activity in multiple myeloma<sup>39</sup>.

### Gene set mutations: NF- $\kappa$ B pathway

Another approach to identify biologically relevant mutations in multiple myeloma is to look not at the frequency of mutation of individual genes, but rather of sets of genes.

We first considered gene sets based on existing insights into the biology of multiple myeloma. For example, activation of the NF- $\kappa$ B pathway is known in multiple myeloma, but the basis of such activation is only partially understood<sup>2,3</sup>. We observed 10 point mutations ( $P = 0.016$ ) and 4 structural rearrangements, affecting 11 NF- $\kappa$ B pathway genes (Supplementary Table 7): *BTRC*, *CARD11*, *CYLD*, *IKBIP*, *IKBKB*, *MAP3K1*, *MAP3K14*, *RIPK4*, *TLR4*, *TNFRSF1A* and *TRAF3*. Taken together, our findings greatly expand the mechanisms by which NF- $\kappa$ B may be activated in multiple myeloma.

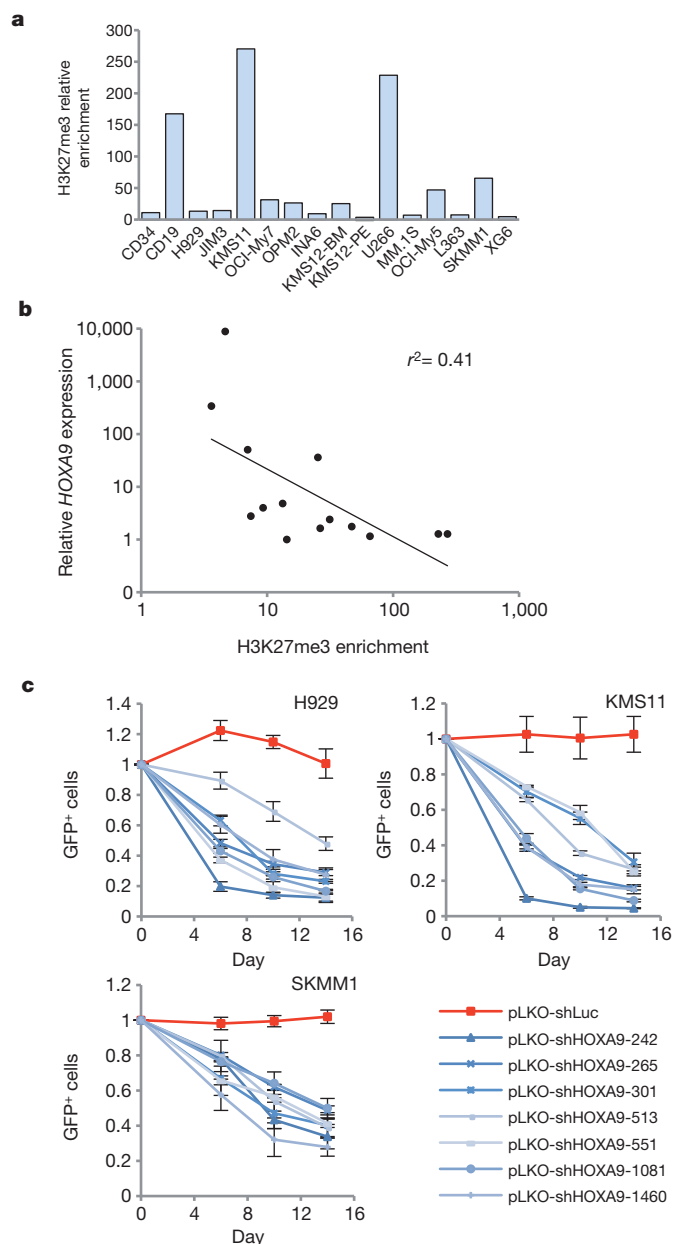
### Gene set mutations: histone modifying enzymes

We next looked for enrichment in mutations in histone-modifying enzymes. This hypothesis arose because of our observation that the homeotic transcription factor *HOXA9* was highly expressed in a subset of multiple myeloma patients, particularly those lacking known IgH translocations (Supplementary Fig. 4a). *HOXA9* expression is regulated primarily by histone methyltransferases (HMT) including members of the MLL family. Sensitive polymerase chain reaction with reverse transcription (RT-PCR) analysis showed that *HOXA9* was in fact ubiquitously expressed in multiple myeloma, with most cases exhibiting biallelic expression consistent with dysregulation via an upstream HMT event (Supplementary Fig. 4b, c). Accordingly, we looked for mutations in genes known to regulate *HOXA9* directly. We found significant enrichment ( $P = 0.0024$ ), with mutations in *MLL*, *MLL2*, *MLL3*, *UTX*, *WHSC1* and *WHSC1L1*.

*HOXA9* is normally silenced by histone 3 lysine 27 trimethylation (H3K27me3) chromatin marks when cells differentiate beyond the haematopoietic stem-cell stage<sup>40,41</sup>. This repressive mark was weak or absent at the *HOXA9* locus in most multiple myeloma cell lines (Fig. 3a). Moreover, there was inverse correlation between H3K27me3 levels and *HOXA9* expression (Fig. 3b), consistent with HMT dysfunction contributing to aberrant *HOXA9* expression.

To establish the functional significance of *HOXA9* expression in multiple myeloma cells, we knocked down its expression with seven shRNAs (Supplementary Fig. 5). In 11 out of 12 multiple myeloma cell lines, *HOXA9*-depleted cells exhibited a competitive disadvantage (Fig. 3c and Supplementary Fig. 6).

These experiments indicate that aberrant *HOXA9* expression, caused at least in part by HMT-related genomic events, has a role



**Figure 3 | *HOXA9* is a candidate oncogene in multiple myeloma.**

**a**, H3K27me3 enrichment at the *HOXA9* promoter in CD34 cells, CD19 cells and multiple myeloma cell lines relative to H3K27me3 methylation at the BC site, known to be hypomethylated in all cells. **b**, Relative *HOXA9* expression against H3K27me3 enrichment at the *HOXA9* locus. **c**, GFP competition assay in multiple myeloma cell lines. After lentiviral infection with seven *HOXA9* shRNAs or a control shRNA targeting luciferase, GFP-positive cells were monitored by flow cytometry and compared to the proportion of GFP-positive cells present in the population 3 days after infection (designated day 0). Error bars indicate standard error of the mean and represent a minimum of three independent experiments.

in multiple myeloma and may represent a new therapeutic target. Further supporting a role of *HOXA9* as a multiple myeloma oncogene, array-based comparative genomic hybridization identified focal amplifications of the *HOXA* locus in 5% of patients (Supplementary Fig. 7).

### Discovering new gene set mutations

We next asked whether it would be possible to discover pathways enriched for mutations in the absence of previous knowledge. Accordingly, we examined 616 gene sets in the MSigDB Canonical



**Table 2 | Statistically significant mutated non-coding regions**

| Chr | Start     | End       | Length (nt) | Mut. | Samples | P-value                | q-value              | Separation (nt)           | Gene                    | Coding events         |
|-----|-----------|-----------|-------------|------|---------|------------------------|----------------------|---------------------------|-------------------------|-----------------------|
| 1   | 554350    | 555310    | 960         | 3    | 3       | $3.86 \times 10^{-6}$  | 0.11                 | 494, 44                   | AK125248 (intron)       | –                     |
| 1   | 82793220  | 82793300  | 80          | 2    | 2       | $8.39 \times 10^{-6}$  | 0.19                 | 8                         | TLL7/LPHN2 (IGR)        | –                     |
| 1   | 147333070 | 147335140 | 2,070       | 4    | 3       | $2.47 \times 10^{-6}$  | 0.09                 | 350, 1, 85                | NBPFA (intron)          | –                     |
| 2   | 40865560  | 40865630  | 70          | 2    | 2       | $4.99 \times 10^{-6}$  | 0.14                 | 2                         | SLC8A1/PKDCC (IGR)      | –                     |
| 3   | 149273920 | 149274010 | 90          | 2    | 2       | $4.80 \times 10^{-6}$  | 0.14                 | 78                        | ZIC4/AGTR1 (IGR)        | –                     |
| 3   | 189142550 | 189143600 | 1,050       | 8    | 5       | $5.55 \times 10^{-14}$ | $3.9 \times 10^{-8}$ | 298, 8, 17, 26, 26, 80, 1 | BCL6/LPP (IGR)          | –                     |
| 3   | 189440810 | 189441310 | 500         | 3    | 3       | $2.64 \times 10^{-6}$  | 0.09                 | 1, 291                    | LPP (intron)            | –                     |
| 4   | 7819430   | 7819530   | 100         | 2    | 2       | $8.01 \times 10^{-6}$  | 0.18                 | 26                        | AFAP1 (intron)          | Missense mutation     |
| 4   | 39875900  | 39876610  | 710         | 3    | 2       | $5.88 \times 10^{-6}$  | 0.16                 | 109, 412                  | RHOH (intron)           | –                     |
| 4   | 62180540  | 62181370  | 830         | 3    | 3       | $1.05 \times 10^{-5}$  | 0.22                 | 211, 432                  | LPHN3 (intron)          | –                     |
| 4   | 157902080 | 157904460 | 2,380       | 4    | 4       | $6.95 \times 10^{-6}$  | 0.17                 | 996, 423, 443             | PDGFC (3' UTR/intron)   | –                     |
| 7   | 92754250  | 92754270  | 20          | 2    | 2       | $2.03 \times 10^{-7}$  | 0.02                 | 1                         | CCDC132 (intron)        | –                     |
| 9   | 16564360  | 16565100  | 740         | 3    | 2       | $8.65 \times 10^{-6}$  | 0.19                 | 250, 76                   | BNC2 (intron)           | –                     |
| 12  | 120943010 | 120943460 | 450         | 3    | 3       | $6.99 \times 10^{-7}$  | 0.04                 | 17, 9                     | BCL7A (promoter)        | –                     |
| 12  | 120943580 | 120946950 | 3,370       | 4    | 3       | $1.47 \times 10^{-8}$  | 0.0017               | 2055, 657, 295            | BCL7A (promoter/intron) | –                     |
| 14  | 68327320  | 68333190  | 5,870       | 4    | 4       | $7.05 \times 10^{-6}$  | 0.17                 | 397, 156, 35              | ZFP36L1 (intron)        | Indel                 |
| 17  | 8106910   | 8111850   | 4,940       | 4    | 2       | $4.85 \times 10^{-6}$  | 0.14                 | 1483, 389, 83             | PFAS (intron)           | Complex rearrangement |
| 20  | 60328960  | 60329510  | 550         | 2    | 2       | $1.42 \times 10^{-6}$  | 0.06                 | 120                       | LAMA5 (intron)          | Missense mutation     |

Regions of predicted regulatory potential showing mutation frequency beyond that expected by chance are shown ( $q < 0.25$ ). Mut., mutations. 'Start' and 'End' columns indicate the first and last nucleotide of regions of regulatory potential according to hg18/NCBI36. 'Separation' column indicates the number of nucleotides within the regulatory region separating the observed mutations.

Pathways database. One top-ranking gene set was of particular interest because it did not relate to genes known to be important in multiple myeloma. This gene set encodes proteins involved in the formation of the fibrin clot in the blood coagulation cascade. There were 6 mutations, in 5 of 38 patients (16%,  $q = 0.0054$ ), encoding 5 proteins (Supplementary Table 8). RT-PCR analysis confirmed expression of 4 of the 5 coagulation factors in multiple myeloma cell lines (Supplementary Fig. 8). The coagulation cascade involves a number of extracellular proteases and their substrates and regulators, but their role in multiple myeloma has not been suspected. However, thrombin and fibrin have been shown to serve as mitogens in other cell types<sup>42</sup>, and have been implicated in metastasis<sup>43</sup>. These observations suggest that coagulation factor mutations should be explored more fully in human cancers.

### Mutations in non-coding regions

Analyses of non-coding portions of the genome have not previously been reported in cancer. We focused on non-coding regions with highest regulatory potential. We defined  $2.4 \times 10^6$  regulatory potential regions (Supplementary Fig. 9), averaging 280 base pairs (bp). We then treated these regions as if they were protein-coding genes, subjecting them to the same permutation analysis used for exonic regions.

We identified multiple non-coding regions with high frequencies of mutation which fell into two classes (Table 2 and Supplementary Table 9). The first corresponds to regions of known somatic hypermutation. These have a 1,000-fold higher than expected mutation frequency, as expected for post-germinal centre B cells (Supplementary Table 9). These regions comprise immunoglobulin-coding genes and the 5' UTR of the lymphoid oncogene, *BCL6*, as reported<sup>44</sup>. Interestingly, we also found previously unrecognized mutations in the intergenic region flanking *BCL6* in five patients, indicating that somatic hypermutation probably occurs in regions beyond the 5' UTR and first intron of *BCL6* (Table 2). Whether such non-coding *BCL6* mutations contribute to multiple myeloma pathogenesis remains to be established.

The second class consisted of 18 non-coding regions with mutation frequencies beyond that expected by chance ( $q < 0.25$ ) (Table 2 and Supplementary Table 10). Four of the 18 regions flanked genes that also harboured coding mutations. Interestingly, we observed 7 mutations in 5 of 23 patients (22%) within non-coding regions of *BCL7A*, a putative tumour suppressor gene discovered in the B-cell malignancy Burkitt lymphoma<sup>45</sup>, and which is also deleted or hypermethylated in cutaneous T-cell lymphomas<sup>46,47</sup>. The function of *BCL7A* is unknown, and the effect of its non-coding mutations in multiple myeloma remains to be established.

Our preliminary analysis of non-coding mutations indicates that non-exonic portions of the genome may represent a previously untapped source of insight into the pathogenesis of cancer.

### Discussion

The analysis of multiple myeloma genomes reveals that mechanisms previously suspected to have a role in the biology of multiple myeloma (for example, NF- $\kappa$ B activation and HMT dysfunction) may have broad roles by virtue of mutations in multiple members of these pathways. In addition, potentially new mechanisms of transformation are suggested, including mutations in the RNA exonuclease *DIS3* and other genes involved in protein translation and homeostasis. Whether these mutations are unique to multiple myeloma or are common to other cancers remains to be determined. Furthermore, frequent mutations in the oncogenic kinase *BRAF* were observed—a finding that has immediate clinical translational implications.

Importantly, most of these discoveries could not have been made by sequencing only a single multiple myeloma genome—the complex patterns of pathway dysregulation required the analysis of multiple genomes. Whole-exome sequencing revealed the substantial majority of the significantly mutated genes. However, we note that half of total protein-coding mutations occurred via chromosomal aberrations such as translocations, most of which would not have been discovered by sequencing of the exome alone. Similarly, the recurrent point mutations in non-coding regions would have been missed with sequencing directed only at coding exons.

The analysis described here is preliminary. Additional multiple myeloma genomes will be required to establish the definitive genomic landscape of the disease and determine accurate estimates of mutation frequency in the disease. The sequence data described here will be available from the dbGaP repository (<http://www.ncbi.nlm.nih.gov/gap>) and we have created a multiple myeloma Genomics Portal (<http://www.broadinstitute.org/mmmp>) to support data analysis and visualization.

### METHODS SUMMARY

Informed consent from multiple myeloma patients was obtained in line with the Declaration of Helsinki. DNA was extracted from bone marrow aspirate (tumour) and blood (normal). WGS libraries (370–410-bp inserts) and WES libraries (200–350-bp inserts) were constructed and sequenced on an Illumina GA-II sequencer using 101- and 76-bp paired-end reads, respectively. Sequencing reads were processed with the Firehose pipeline, identifying somatic point mutations, indels and other structural chromosomal rearrangements. Structural rearrangements affecting protein-coding regions were then subjected to manual review to exclude alignment artefacts. True positive mutation rates were estimated by Sequenom mass spectrometry genotyping of randomly selected mutations. *HOXA9* short hairpin

(sh)RNAs were introduced into multiple myeloma cell lines using lentiviral infection using standard methods.

A complete description of the materials and methods is provided in the Supplementary Information.

Received 11 June 2010; accepted 17 January 2011.

- Bergsagel, P. L. & Kuehl, W. M. Molecular pathogenesis and a consequent classification of multiple myeloma. *J. Clin. Oncol.* **23**, 6333–6338 (2005).
- Keats, J. J. *et al.* Promiscuous mutations activate the noncanonical NF- $\kappa$ B pathway in multiple myeloma. *Cancer Cell* **12**, 131–144 (2007).
- Annunziata, C. M. *et al.* Frequent engagement of the classical and alternative NF- $\kappa$ B pathways by diverse genetic abnormalities in multiple myeloma. *Cancer Cell* **12**, 115–130 (2007).
- van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41**, 521–523 (2009).
- Lee, W. *et al.* The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465**, 473–477 (2010).
- Campbell, P. J. *et al.* Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nature Genet.* **40**, 722–729 (2008).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Ding, L. *et al.* Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**, 999–1005 (2010).
- Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
- Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Reva, B., Antipin, Y. & Sander, C. Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol.* **8**, R232 (2007).
- Dziembowski, A. *et al.* A single subunit, Dis3, is essentially responsible for yeast exosome core activity. *Nature Struct. Mol. Biol.* **14**, 15–22 (2007).
- Schmid, M. & Jensen, T. H. The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* **33**, 501–510 (2008).
- Schneider, C., Anderson, J. T. & Tollervey, D. The exosome subunit Rps44 plays a direct role in RNA substrate recognition. *Mol. Cell* **27**, 324–331 (2007).
- Barbas, A. *et al.* Determination of key residues for catalysis and RNA cleavage specificity: one mutation turns RNase II into a “SUPER-ENZYME”. *J. Biol. Chem.* **284**, 20486–20498 (2009).
- Ibrahim, H., Wilusz, J. & Wilusz, C. J. RNA recognition by 3′-to-5′ exonucleases: the substrate perspective. *Biochim. Biophys. Acta* **1779**, 256–265 (2008).
- Zhan, F. *et al.* The molecular classification of multiple myeloma. *Blood* **108**, 2020–2028 (2006).
- Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
- Mootha, V. K. *et al.* PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
- Tanay, A., Regev, A. & Shamir, R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl Acad. Sci. USA* **102**, 7203–7208 (2005).
- Carrasco, D. R. *et al.* The differentiation and stress response factor XBP-1 drives multiple myeloma pathogenesis. *Cancer Cell* **11**, 349–360 (2007).
- Zimprich, A. *et al.* Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601–607 (2004).
- Paisán-Ruiz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**, 595–600 (2004).
- Forman, M. S., Lee, V. M. & Trojanowski, J. Q. ‘Unfolding’ pathways in neurodegenerative disease. *Trends Neurosci.* **26**, 407–410 (2003).
- Masciarelli, S. *et al.* CHOP-independent apoptosis and pathway-selective induction of the UPR in developing plasma cells. *Mol. Immunol.* **47**, 1356–1365 (2010).
- Cenci, S. & Sitia, R. Managing and exploiting stress in the antibody factory. *FEBS Lett.* **581**, 3652–3657 (2007).
- Todd, D. J., Lee, A. H. & Glimcher, L. H. The endoplasmic reticulum stress response in immunity and autoimmunity. *Nature Rev. Immunol.* **8**, 663–674 (2008).
- Adams, J. The development of proteasome inhibitors as anticancer drugs. *Cancer Cell* **5**, 417–421 (2004).
- Shaffer, A. L. *et al.* IRF4 addiction in multiple myeloma. *Nature* **454**, 226–231 (2008).
- Mandelbaum, J. *et al.* BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell* **18**, 568–579 (2010).
- Pasqualucci, L. *et al.* Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J. Exp. Med.* **203**, 311–317 (2006).
- Shaffer, A. L. *et al.* Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51–62 (2002).
- Shapiro-Shelef, M. *et al.* Blimp-1 is required for the formation of immunoglobulin secreting plasma cells and pre-plasma memory B cells. *Immunity* **19**, 607–620 (2003).
- Turner, C. A. Jr, Mack, D. H. & Davis, M. M. Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell* **77**, 297–306 (1994).
- Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Flaherty, K. *et al.* Phase I study of PLX4032: Proof of concept for V600E *BRAF* mutation as a therapeutic target in human cancer. In 2009 ASCO Meeting (American Society of Clinical Oncology, 2009).
- Kim, K. *et al.* Blockade of the MEK/ERK signalling cascade by AS703026, a novel selective MEK1/2 inhibitor, induces pleiotropic anti-myeloma activity *in vitro* and *in vivo*. *Br. J. Haematol.* **149**, 537–549 (2010).
- Lessard, J. & Sauvageau, G. Polycomb group genes as epigenetic regulators of normal and leukemic hematopoiesis. *Exp. Hematol.* **31**, 567–585 (2003).
- Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
- Ruf, W. & Mueller, B. M. Thrombin generation and the pathogenesis of cancer. *Semin. Thromb. Hemost.* **32** (suppl. 1), 61–68 (2006).
- Esumi, N., Fan, D. & Fidler, I. J. Inhibition of murine melanoma experimental metastasis by recombinant desulfatohirudin, a highly specific thrombin inhibitor. *Cancer Res.* **51**, 4549–4556 (1991).
- Migliozzi, A. *et al.* Frequent somatic hypermutation of the 5′ noncoding region of the *BCL6* gene in B-cell lymphoma. *Proc. Natl Acad. Sci. USA* **92**, 12520–12524 (1995).
- Zani, V. J. *et al.* Molecular cloning of complex chromosomal translocation t(8;14)(q24.1;q32.3;q24.1) in a Burkitt lymphoma cell line defines a new gene (*BCL7A*) with homology to caldesmon. *Blood* **87**, 3124–3134 (1996).
- Zhang, W. *et al.* Unravelling the hidden heterogeneities of diffuse large B-cell lymphoma based on coupled two-way clustering. *BMC Genomics* **8**, 332 (2007).
- Carbone, A. *et al.* Array-based comparative genomic hybridization in early-stage mycosis fungoides: recurrent deletion of tumor suppressor genes *BCL7A*, *SMAC/DIABLO*, and *RHOA*. *Genes Chromosom. Cancer* **47**, 1067–1075 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This project was funded by a grant from the Multiple Myeloma Research Foundation. M.A.C. was supported by a Clinician Scientist Fellowship from Leukaemia and Lymphoma Research (UK). We are grateful to all members of the Broad Institute's Biological Samples Platform, Genetic Analysis Platform, and Genome Sequencing Platform, without whom this work would not have been possible.

**Author Contributions** M.A.C., J.J.K., A.C.S., C.L.H., M.A. and B.E.B. performed experimental work, including PCR, cloning, ChIP analyses and RNAi experiments. M.A.C., M.S.L., J.J.K., K.C., J.-P.B., Y.D., S.M., T.J.P., A.H.R., A.S., D.V. and G.G. performed data analyses. M.A.C., M.S.L., K.C., E.S.L., G.G. and T.R.G. produced the text and figures, including Supplementary Information. J.C., J.T., W.C.H., L.A.G., M.M., E.S.L., G.G. and T.R.G. provided leadership for the project. All authors contributed to the final manuscript. K.C.A., R.F., C.C.H., S.J., A.J.J., A.K., T.L., S.L., S.V.R., D.S.S., A.K.S., S.T., R.V. and T.Z. collected data and provided patient materials. J.J.K., C.S., G.J.A., K.G.A., D.A., A.B., P.L.B., S.B.G., J.L., T.L., S.M., B.M., L.M.P., R.O., W.W. and J.C. processed and analysed genetic material, including RNA/DNA extraction, fingerprinting, genotyping, data management, hybridizations, library preparation and sequencing.

**Author Information** Sequence data have been deposited in the dbGaP repository (<http://www.ncbi.nlm.nih.gov/gap>) under accession number phs000348.v1.p1. Additional data have been submitted to the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo>) under accession numbers GSE26760 (GEP data) and GSE26849 (aCGH data); both data sets are also combined under accession code GSE26863. We have also created a multiple myeloma Genomics Portal (<http://www.broadinstitute.org/mmgp>) to support data analysis and visualization. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.R.G. ([golub@broadinstitute.org](mailto:golub@broadinstitute.org)) or G.G. ([gadgets@broadinstitute.org](mailto:gadgets@broadinstitute.org)).



# The developmental transcriptome of *Drosophila melanogaster*

Brenton R. Graveley<sup>1\*</sup>, Angela N. Brooks<sup>2\*</sup>, Joseph W. Carlson<sup>3\*</sup>, Michael O. Duff<sup>1\*</sup>, Jane M. Landolin<sup>3\*</sup>, Li Yang<sup>1\*</sup>, Carlo G. Artieri<sup>4</sup>, Marijke J. van Baren<sup>5</sup>, Nathan Boley<sup>6</sup>, Benjamin W. Booth<sup>3</sup>, James B. Brown<sup>6</sup>, Lucy Cherbas<sup>7</sup>, Carrie A. Davis<sup>8</sup>, Alex Dobin<sup>8</sup>, Renhua Li<sup>4</sup>, Wei Lin<sup>8</sup>, John H. Malone<sup>4</sup>, Nicolas R. Mattiuzzo<sup>4</sup>, David Miller<sup>9</sup>, David Sturgill<sup>4</sup>, Brian B. Tuch<sup>10,11</sup>, Chris Zaleski<sup>8</sup>, Dayu Zhang<sup>7</sup>, Marco Blanchette<sup>12,13</sup>, Sandrine Dudoit<sup>14</sup>, Brian Eads<sup>9</sup>, Richard E. Green<sup>15</sup>, Ann Hammonds<sup>3</sup>, Lichun Jiang<sup>4</sup>, Phil Kapranov<sup>8</sup>, Laura Langton<sup>5</sup>, Norbert Perrimon<sup>16</sup>, Jeremy E. Sandler<sup>3</sup>, Kenneth H. Wan<sup>3</sup>, Aaron Willingham<sup>17</sup>, Yu Zhang<sup>4</sup>, Yi Zou<sup>7</sup>, Justen Andrews<sup>9</sup>, Peter J. Bickel<sup>6</sup>, Steven E. Brenner<sup>2,17</sup>, Michael R. Brent<sup>5</sup>, Peter Cherbas<sup>7,9</sup>, Thomas R. Gingeras<sup>8,18</sup>, Roger A. Hoskins<sup>3</sup>, Thomas C. Kaufman<sup>9</sup>, Brian Oliver<sup>4</sup> & Susan E. Celniker<sup>3</sup>

*Drosophila melanogaster* is one of the most well studied genetic model organisms; nonetheless, its genome still contains unannotated coding and non-coding genes, transcripts, exons and RNA editing sites. Full discovery and annotation are pre-requisites for understanding how the regulation of transcription, splicing and RNA editing directs the development of this complex organism. Here we used RNA-Seq, tiling microarrays and cDNA sequencing to explore the transcriptome in 30 distinct developmental stages. We identified 111,195 new elements, including thousands of genes, coding and non-coding transcripts, exons, splicing and editing events, and inferred protein isoforms that previously eluded discovery using established experimental, prediction and conservation-based approaches. These data substantially expand the number of known transcribed elements in the *Drosophila* genome and provide a high-resolution view of transcriptome dynamics throughout development.

*Drosophila melanogaster* is an important non-mammalian model system that has had a critical role in basic biological discoveries, such as identifying chromosomes as the carriers of genetic information<sup>1</sup> and uncovering the role of genes in development<sup>2,3</sup>. Because it shares a substantial genic content with humans<sup>4</sup>, *Drosophila* is increasingly used as a translational model for human development, homeostasis and disease<sup>5</sup>.

High-quality maps are needed for all functional genomic elements. Previous studies demonstrated that a rich collection of genes is deployed during the life cycle of the fly<sup>6–8</sup>. Although expression profiling using microarrays has revealed the expression of ~13,000 annotated genes, it is difficult to map splice junctions and individual base modifications generated by RNA editing<sup>9</sup> using such approaches. Single-base resolution is essential to define precisely the elements that comprise the *Drosophila* transcriptome.

Estimates of the number of transcript isoforms are less accurate than estimates of the number of genes. Whereas ~20% of *Drosophila* genes are annotated as encoding alternatively spliced pre-mRNAs, splice-junction microarray experiments indicate that this number is at least 40% (ref. 7). Determining the diversity of mRNAs generated by alternative promoters, alternative splicing and RNA editing will substantially increase the inferred protein repertoire. Non-coding RNA genes (ncRNAs) including short interfering RNAs (siRNAs) and

microRNAs (miRNAs) (reviewed in ref. 10), and longer ncRNAs such as *bxd* (ref. 11) and *rox* (ref. 12), have important roles in gene regulation, whereas others such as small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs) are important components of macromolecular machines such as the ribosome and spliceosome. The transcription and processing of these ncRNAs must also be fully documented and mapped.

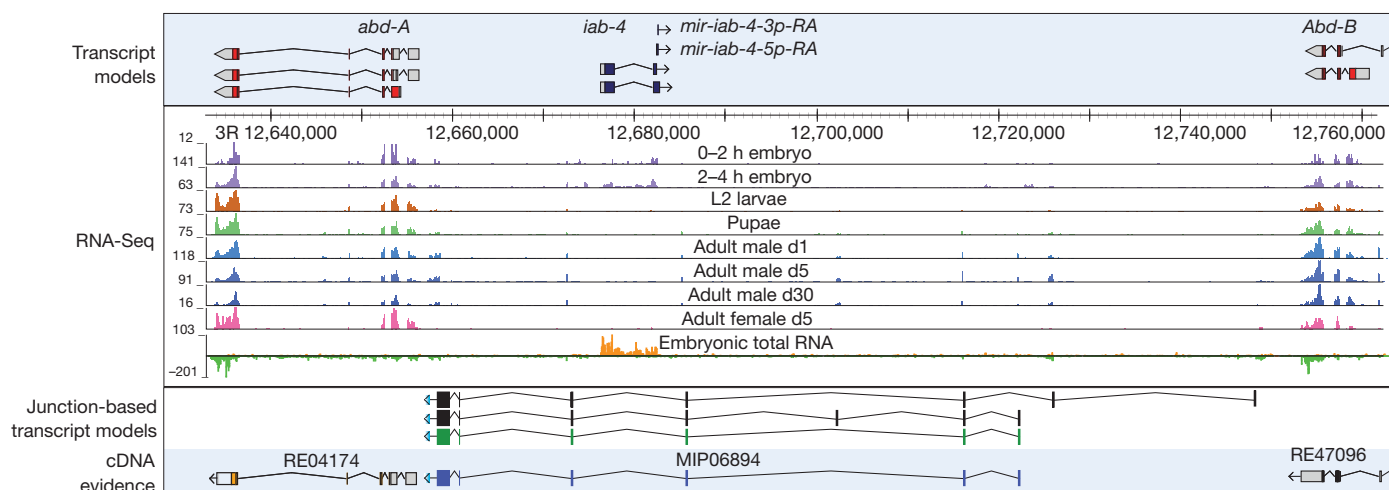
As part of the modENCODE project to annotate the functional elements of the *D. melanogaster* and *Caenorhabditis elegans* genomes<sup>13–15</sup>, we used RNA-Seq and tiling microarrays to sample the *Drosophila* transcriptome at unprecedented depth throughout development from early embryo to ageing male and female adults. We report on a high-resolution view of the discovery, structure and dynamic expression of the *D. melanogaster* transcriptome.

## Strategy for characterization of the transcriptome

To discover new transcribed features (Supplementary Table 1) and comprehensively characterize their expression dynamics throughout development, we conducted complementary tiling microarray and RNA-Seq experiments using RNA isolated from 30 whole-animal samples representing 27 distinct stages of development (Supplementary Table 2). These included 12 embryonic samples collected at 2-h intervals for 24 h, six larval, six pupal and three sexed adult stages at 1, 5

<sup>1</sup>Department of Genetics and Developmental Biology, University of Connecticut Health Center, 263 Farmington Avenue, Farmington, Connecticut 06030-6403, USA. <sup>2</sup>Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA. <sup>3</sup>Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>4</sup>Section of Developmental Genomics, Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. <sup>5</sup>Center for Genome Sciences and Department of Computer Science, Washington University, St. Louis, Missouri 63108, USA. <sup>6</sup>Department of Statistics, University of California, Berkeley, California, 94720 USA. <sup>7</sup>Center for Genomics and Bioinformatics, Indiana University, 1001 E. 3rd Street, Bloomington, Indiana 47405-7005, USA. <sup>8</sup>Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. <sup>9</sup>Department of Biology, Indiana University, 1001 E. 3rd Street, Bloomington, Indiana 47405-7005, USA. <sup>10</sup>Genetic Systems Division, Research and Development, Life Technologies, Foster City, California 94404, USA. <sup>11</sup>Genome Analysis Unit, Amgen, South San Francisco, California 94080, USA. <sup>12</sup>Stowers Institute for Medical Research, 1000 East 50th street, Kansas City, Missouri 64110, USA. <sup>13</sup>Department of Pathology and Laboratory Medicine, Kansas University Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA. <sup>14</sup>Division of Biostatistics, School of Public Health, University of California, Berkeley, California 94720, USA. <sup>15</sup>Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California 95064, USA. <sup>16</sup>Department of Genetics and Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>17</sup>Department of Plant and Microbial Biology, University of California, Berkeley, California 94720, USA. <sup>18</sup>Affymetrix, Santa Clara, California 95051, USA.

\*These authors contributed equally to this work.



**Figure 1 | Discovery of new RNAs in the Bithorax complex.** Genomic organization and experimental evidence for new transcripts located between the HOX genes, *abd-A* and *Abd-B*, based on short poly(A)<sup>+</sup> RNA and total RNA-Seq expression profiles. The numbers to the left of each track indicate the

maximal number of reads for that sample. Three manually curated junction-based transcript models are shown; the green transcript model was fully validated by a cDNA, MIP06894.

and 30 days after eclosion. We used 38-base-pair (bp) resolution genome tiling microarrays to analyse total RNA from all 30 biological samples and poly(A)<sup>+</sup> mRNA from the 12 embryonic samples (Supplementary Fig. 1). To attain single-nucleotide resolution and to facilitate the analysis of alternative splicing and RNA editing, we performed non-strand-specific poly(A)<sup>+</sup> RNA-Seq from all 30 samples generating a combination of single and paired-end ~75-bp reads on the Illumina Genome Analyser IIX platform (short poly(A)<sup>+</sup> RNA-Seq) (Supplementary Table 3 and Supplementary Fig. 2). To identify primary transcripts and non-coding RNAs, the 12 embryonic time points were also interrogated with strand-specific 50-bp sequence reads from partially rRNA-depleted total RNA on the Applied Biosystems SOLiD platform (Supplementary Table 4 and Supplementary Fig. 3). To improve connectivity, mixed-stage embryos, adult males and adult females were used to generate ~250-bp reads on the Roche 454 platform (non-strand-specific long poly(A)<sup>+</sup> RNA-Seq) (Supplementary Table 5). In total, we generated 176,962,906,041 bp of mapped sequence representing 1,266-fold coverage of the genome and 5,902-fold coverage of the annotated *D. melanogaster* transcriptome.

### Discovery of new transcribed regions

We identified 1,938 new transcribed regions (NTRs) not linked to any annotated gene models. Herein, 'transcripts' refer to RNA molecules synthesized from a genomic locus whereas 'genes' refer to one or more transcripts that share exons in their mature spliced form. modENCODE cDNAs fully support 13% of the NTRs (Supplementary Fig. 4) and partially support 23%. Most NTRs (84%) are detected by poly(A)<sup>+</sup> RNA-Seq, 44% by total RNA-Seq and 42% by tiling array. Approximately half of the NTRs are conserved in the distantly related *Drosophila pseudoobscura* and *Drosophila mojavensis* (Supplementary Fig. 4b) and 30% of these are detected by poly(A)<sup>+</sup> RNA-Seq data from *D. pseudoobscura* or *D. mojavensis* adult heads (Supplementary Fig. 4c, d, Supplementary Table 6 and Supplementary Methods). The NTRs probably eluded previous detection because they are expressed at low levels, in temporally restricted patterns, and are enriched for single-exon genes. The new multi-exon gene models (48%) have fewer, shorter and less conserved exons than annotated genes.

Nearly one-third of the NTRs have a predicted open reading frame (ORF) greater than 100 amino acids. The remaining NTRs could encode small peptides but many are likely to be non-coding RNAs. A small fraction (9%) of NTRs are heterochromatic; most of these (232) have sequence similarity (greater than 100-nucleotide match and greater than 60% identity) to transposable elements (TEs) and

represent transcribed TEs or TE fragments. It remains to be determined whether these regions have any function, although recent studies describe TE-associated regions that have acquired functions<sup>16,17</sup>.

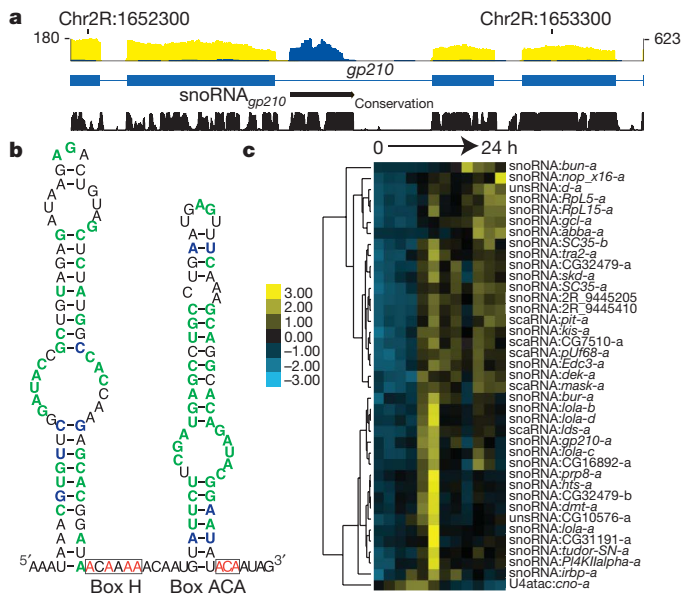
Even in the well-studied Bithorax complex<sup>2</sup> we found an NTR. Known genetic breakpoints in the infra-abdominal regions *iab-3* to *iab-8*, which lie between the homeotic genes *abdominal A* (*abd-A*) and *Abdominal B* (*Abd-B*), disrupt normal male development and affect fertility<sup>18,19</sup>. Within this region are regulatory elements<sup>20</sup> and evidence for long non-coding RNAs that have eluded detection for over 20 years<sup>21–23</sup>. We used the RNA-Seq data to infer the structures of at least three overlapping transcripts and validated one form (Fig. 1). The RNAs are expressed in embryos and adult males but not females. On the basis of the presumed role of this new gene and spatial expression in the embryonic gonad (data not shown), we have named it *male specific abdominal* (*msa*). The cDNA contains short ORFs that are conserved in the *melanogaster* subgroup and could encode male-specific peptides. Whether they function as regulatory and/or as peptide-encoding RNAs is an important question for understanding development and segmental morphological diversity.

### Discovery of small ncRNAs

We identified 37 unannotated intron-encoded and two unannotated intergenic small ncRNAs (<300 nucleotides) with an average fragments per kilobase of transcript per million fragments mapped (FPKM)<sup>24</sup> >20 from total embryonic RNA-Seq (Fig. 2 and Supplementary Table 7). Most of these ncRNAs are highly conserved in *Drosophila* sibling species<sup>25</sup>. We found published but unannotated ncRNAs: a U4atac snRNA<sup>26</sup> and four small Cajal-body-specific RNAs (scaRNAs)<sup>27</sup>. Of the remaining 34 ncRNAs, three are box C/D-like snoRNAs, 28 are box H/ACA-like snRNAs, one is a scaRNA-like RNA, and two are unclassified. One-third of these are located in the introns of genes encoding RNA-binding proteins, the majority of which are involved in pre-mRNA splicing (*xl6*, *SC35*, *tra2*, *Dek*, *Prp8*, *Tudor-SN*, and *pUf68*).

### Discovery of microRNA primary transcripts

MicroRNAs are processed from primary microRNA transcripts (pri-miRNAs) and are either independently transcribed or embedded in the introns of protein-coding genes. We identified 23 putative independently transcribed pri-miRNAs from the total embryonic RNA-Seq and tiling array data that encode 37 annotated miRNAs (Supplementary Table 8). Only two primary transcripts were previously annotated (*bft* and *iab-4*). The pri-miRNAs range from 1 to 18 kb and terminate



**Figure 2 | Discovery of small non-coding RNAs.** **a**, Poly(A)<sup>+</sup> (yellow) and total RNA (blue) data from 10–12-h embryos are shown for the *gp210* gene which hosts a representative new snoRNA. The maximal number of reads in the poly(A)<sup>+</sup> and total RNA-Seq data are shown on the left and right of the track, respectively. **b**, The predicted RNA secondary structure of snoRNA<sub>*gp210*</sub> is characteristic of a H/ACA-box snoRNA. Nucleotides that are 100% conserved in sequence or base-pairing are indicated in green and blue, respectively. **c**, Embryonic expression of the new small RNAs. The scale bar indicates FPKM Z-scores. unsRNA, unclassified small RNA.

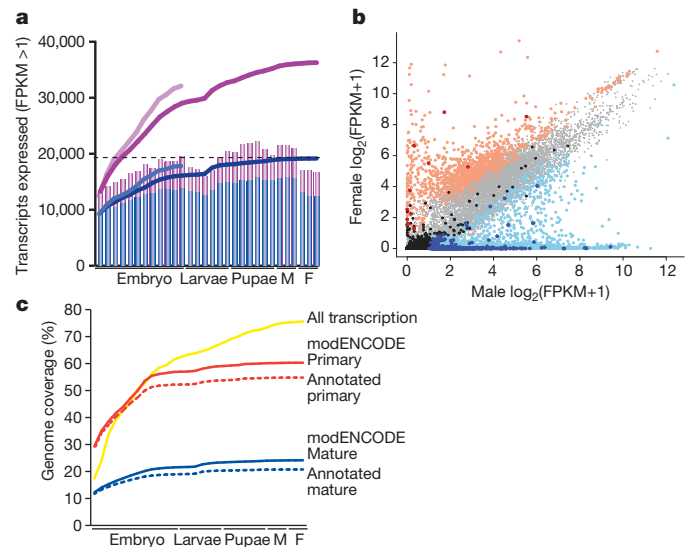
at the mature miRNA (*pre-mir-315*, Supplementary Fig. 5a). Twelve of the 23 precursors have cap analysis of gene expression (CAGE) peaks that map at their initiation sites<sup>28</sup>. pri-miRNA expression is dynamic in embryonic development (Supplementary Fig. 5b).

## Overview of the *Drosophila* transcriptome

We calculated expression levels of annotated genes, transcripts and NTRs (Supplementary Table 9) in the short poly(A)<sup>+</sup> RNA-Seq and tiling array data sets. From the RNA-Seq data we detected expression of 14,862 genes (Supplementary Fig. 7a) and 36,274 transcripts (Fig. 3a) with an FPKM >1 (Supplementary Tables 9–18) of which 67% of genes and 58% of transcripts were also observed in the array data (score >300) (Supplementary Fig. 6 and Supplementary Tables 19 and 20). This includes the confirmation of 87% of annotated genes and transcripts and the discovery of 17,745 new transcripts. In addition, from the total RNA-Seq data we detected expression of 12,854 genes and 32,139 transcripts with an FPKM >1 (Supplementary Tables 12, 13, 21 and 22) of which 77% of genes and 89% of transcripts were also observed in the array data. Of the genes and transcripts observed exclusively in the total RNA-Seq data, 519 genes and 1,005 transcripts (primarily noncoding) were previously annotated and 122 genes and 1,422 transcripts are new discoveries. The genes and transcripts not detected in any data set include small genes (<200 bp), members of multi-copy gene families such as ribosomal RNAs, paralogues (expected owing to our mapping parameters), genes known to be expressed at low levels or in small numbers of cells (for example, gustatory and odorant receptor genes), and non-polyadenylated transcripts.

## Expression dynamics

We examined the dynamics of gene expression throughout development using the short poly(A)<sup>+</sup> RNA-Seq data. The numbers of expressed genes (FPKM >1) (Supplementary Fig. 7a) and transcripts (Fig. 3a) gradually increases, from 7,045 (0–2 h embryos) to 12,000 (adult males). Adult males express ~3,000 more genes than adult females, consistent



**Figure 3 | Dynamics of gene expression.** **a**, Transcripts expressed (FPKM >1) in the short poly(A)<sup>+</sup> RNA-Seq data: FlyBase 5.12, blue; modENCODE, purple. The bar graphs indicate the number of transcripts expressed in each sample (Supplementary Table 1); the lines indicate the cumulative number of expressed transcripts. The lighter blue and purple lines indicate the cumulative number of transcripts expressed in the embryonic total RNA-Seq samples. The horizontal dotted lines indicate the number of expressed previously annotated transcripts. F, female; M, male. **b**, Scatter plot of sex-biased gene expression. Light red, female-biased annotated ( $n = 960$ ); dark red, female-biased NTRs ( $n = 12$ ); light blue, male-biased annotated ( $n = 2,401$ ); dark blue, male-biased NTRs ( $n = 431$ ); light grey, unbiased annotated ( $n = 8,217$ ); black, unbiased NTRs ( $n = 136$ ). **c**, Genome coverage. For each developmental sample, the short poly(A)<sup>+</sup> reads were used to estimate the percentage of the genome covered using a cutoff of two reads. The mature and primary transcripts were inferred for the previously FlyBase 5.12 (dotted lines) and modENCODE (solid lines) gene models.

with the known transcriptional complexity of the testis<sup>29</sup>. We observed that 40% of expressed genes are constitutively expressed in 30 samples (Supplementary Fig. 7b). We also observed developmentally regulated expression of TEs (Supplementary Materials and Supplementary Fig. 8).

We observed pronounced expression changes in over 1,500 genes in the first two third instar larval samples (Supplementary Fig. 7a, c). Expression of 1,199 genes increased at least tenfold, and 421 genes decreased at least tenfold (Supplementary Table 23). Nearly all of the upregulated genes are expressed for the first time during the third instar stage and most are poorly characterized genes.

The earliest known event in metamorphosis is the ‘mid-3rd transition’<sup>30</sup>, identified by the synchronous changes in the transcription of a number of well studied genes, *Ecdysone-induced protein 28/29kD* and *Fat body protein 1* (reviewed in ref. 31), and the switch from proximal to distal promoters of *Alcohol dehydrogenase*<sup>32</sup>. These markers coincide with the surge reported here. The mid-3rd transition has no morphological or behavioural correlates and is associated with a pulse of the steroid hormone ecdysone<sup>33</sup> acting through a non-standard receptor<sup>34</sup>. Whether the onset of testis development is a consequence of the mid-3rd transition, or whether the two events are functionally related, remains to be investigated.

Over 29% of protein-coding genes showed significant sex-biased expression in adults (false discovery rate <0.1%), with more male-biased (1,829) or male-specific genes (572) than female-biased (945) or female-specific genes (15) (Supplementary Tables 24 and 25, and Fig. 3b). Known female (*ovo* and *otu*) and male (*dj*) sex-biased genes were expressed as expected. We found that 74% of the NTRs expressed in adults were significantly male-biased whereas only 2.1% were significantly female-biased.



**Table 1 | Classification of alternative splicing events**

| Splicing event               | Diagram | FlyBase r5.12 | modENCODE       | New events      | Short poly(A) <sup>+</sup> RNA-Seq | Significantly changing |
|------------------------------|---------|---------------|-----------------|-----------------|------------------------------------|------------------------|
| Cassette exons               |         | 793           | 2,717           | 2,014           | 2,369                              | 1,539                  |
| Alternative 5' splice sites  |         | 843           | 5,192           | 4,599           | 4,583                              | 3,142                  |
| Alternative 3' splice sites  |         | 879           | 6,253           | 5,505           | 5,579                              | 3,242                  |
| Mutually exclusive exons     |         | 229           | 251             | 123             | 228                                | 226                    |
| Coordinate cassette exons    |         | 301           | 1,227           | 979             | 992                                | 467                    |
| Alternative first exons      |         | 1,767         | 4,936           | 3,442           | 4,473                              | 3,996                  |
| Alternative last exons       |         | 227           | 604             | 432             | 553                                | 471                    |
| Retained/unprocessed introns |         | 1,434         | 2,679 (5,667)   | 1,275 (4,263)   | 2,439 (35,641)                     | 868 (8,998)            |
| Total                        |         | 6,437         | 23,859 (26,847) | 18,369 (21,478) | 21,216 (54,418)                    | 13,951 (22,081)        |

The number of retained/unprocessed introns in parentheses indicates the total number identified, whereas the number not in parentheses indicates the subset of identified events that have been validated by cDNA sequences or FlyBase 5.12 annotations.

## Genome coverage

Mature mRNAs are encoded by 20% of the *D. melanogaster* genome and primary transcripts by 60% (Fig. 3c). An additional 15% of the genome (~75% total) is detected when considering all of the short poly(A)<sup>+</sup> RNA-Seq data. However, as greater than 99% of the reads map within the bounds of the transcript models, the reads that map to intergenic regions constitute a small minority of our data. Thus, although pervasive transcription of mammalian genomes has been observed in microarray studies<sup>35</sup>, we found little evidence of such 'dark matter'<sup>36</sup> (that is, pervasive transcription) in *D. melanogaster*.

## Discovery and dynamics of alternative splicing

To characterize constitutive and alternative splicing, we identified 71,316 splice junctions, of which 22,965 were new discoveries. Of the new splice junctions, 26% were supported by multiple experimental data types and 74% by only one data type, (Supplementary Fig. 9a) primarily short poly(A)<sup>+</sup> RNA-Seq. Of the 20,751 new junctions from the short poly(A)<sup>+</sup> RNA-Seq data, 7,833 were incorporated into new transcript models or transcribed regions (NTRs). The remaining new junctions have yet to be incorporated into transcript models.

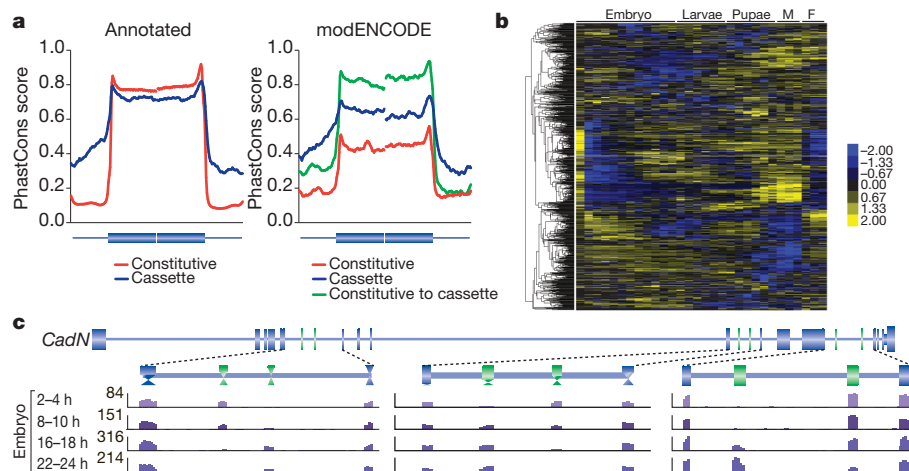
We also identified a total of 102,026 exons (Supplementary Table 26). Of the 52,914 representing new and revised exons, 65% were validated by capture and sequencing of cDNAs and 2,586 were supported by RNA-Seq data from *D. mojavensis* and *D. pseudoobscura*. Of the new exons, 3,392 were identified from the new splice junctions but have yet to be incorporated into transcript models.

To examine splicing dynamics throughout development, we categorized all splicing events into the common types of alternative splicing events (Table 1). We identified a total of 23,859 splicing events, of which 18,369 were new or recategorized, a threefold increase from

annotated splicing events. An additional 2,988 retained/unprocessed introns were identified that were supported by only one experimental data type. In all, 7,473 genes contain at least one alternative splicing event, which is 60.7% of the 12,295 expressed multi-exon genes—also a threefold increase in the fraction of genes with alternatively spliced transcripts. Although smaller than the fraction of human genes with alternatively spliced transcripts (95%)<sup>37,38</sup>, a larger proportion of *Drosophila* genes encode alternative transcripts than was previously known.

Of the new alternative exons, 8,226 were previously annotated as constitutive. As observed<sup>39</sup>, annotated cassette exons, and their flanking introns, are more highly conserved than annotated constitutive exons (Fig. 4a). The newly discovered cassette exons are more highly conserved than the new constitutive exons, although both classes are less conserved than the corresponding class of annotated exons. New cassette exons that were previously annotated as constitutive exons are the most highly conserved set of exons (Fig. 4a). Annotated and new cassette exons show a strong tendency to preserve reading frame (Supplementary Fig. 9b), indicating that these transcripts increase protein diversity. Both annotated and new cassette exons tend to be shorter than their constitutive counterparts, although both sets of new exons tend to be shorter than annotated exons.

To assess the extent of splicing variation we calculated the 'per cent spliced in' or  $\Psi$  (ref. 38) for each splicing event in each sample as well as the switch score ( $\Delta\Psi$ ) by determining the difference between the highest and lowest  $\Psi$  values across development ( $\Delta\Psi = \Psi_{\max} - \Psi_{\min}$ ). This revealed a very smooth distribution of  $\Delta\Psi$  among all events, indicating that the splicing of most exons is fairly constant whereas only a minority change markedly (Supplementary Fig. 9c and Supplementary Table 27). Only 831 splicing events have a  $\Delta\Psi$  value >90. Further statistical



**Figure 4 | Developmentally regulated splicing events.** a, Conservation of internal constitutive and cassette exons >50 nucleotides that were annotated or new discoveries. (Annotated constitutive,  $n = 26,127$ ; annotated cassette,  $n = 438$ ; modENCODE cassette,  $n = 173$ ; modENCODE constitutive,  $n = 306$ ; FlyBase 5.12 constitutive to modENCODE cassette,  $n = 304$ .) b, Clusters of

regulated cassette exon events during development. The scale bar indicates Z-scores of  $\Psi$ . c, Regulated alternative splicing in *CadN* during embryogenesis. The maximal number of reads in the poly(A)<sup>+</sup> RNA-Seq data are indicated for each track.

analyses (see Supplementary Methods) identified 13,951 (66%) alternative splicing events that change significantly throughout development (Supplementary Table 28).

Hierarchical clustering of cassette exon events revealed the dynamic nature of splicing throughout development (Fig. 4b), as exemplified by *Cadherin-N* (*CadN*), a gene with three sets of mutually exclusive exons (Fig. 4c). In each set, one exon is preferentially included in early embryos, the other in late embryos, with a smooth transition between the two. Our analysis also identified groups of exons that have coordinated splicing patterns (Fig. 4b). A set of 55 genes contain exons that are preferentially included in early embryos, late larvae, early pupae and females but skipped in all other stages. Gene Ontology (GO) analysis of these genes indicates that many encode proteins involved in epithelial cell-to-cell junctions. GO analysis of genes that contain exons preferentially included during late pupal and adult stages indicates that many encode proteins that are part of neuronal synapses.

### Sex-biased alternative splicing

Sex determination in *Drosophila* is mediated by a cascade of regulated alternative splicing events involving *Sex lethal* (*Sxl*), *transformer* (*tra*), *male-specific lethal 2* (*msl-2*), *doublesex* (*dsx*) and *fruitless* (*fru*) that specify nearly all physical and behavioural dimorphisms between males and females as well as X chromosome dosage compensation<sup>40</sup>. Our RNA-Seq data confirm sex-biased ( $\Delta\Psi = |\Psi_{\text{male}} - \Psi_{\text{female}}|$ ) splicing of *Sxl* ( $\Delta\Psi = 89.6$ ), *tra* ( $\Delta\Psi = 39.2$ ), *dsx* ( $\Delta\Psi = 59.7$ ) and *fru* ( $\Delta\Psi = 100$ ).

In addition to the canonical sex-determination cascade, we identified 119 strongly sex-biased splicing events ( $\Delta\Psi > 70$ ) (Supplementary Fig. 9d). One striking example is *Reps*, which was annotated as containing six constitutive exons. RNA-Seq data indicate that exon five is a sex-biased alternative cassette exon ( $\Delta\Psi = 73.39$ ) (Supplementary Fig. 10). This highly conserved exon is included in males and skipped in females. The intron upstream of this cassette exon contains conserved SXL binding sites, indicating that it is regulated by SXL and is a candidate sex differentiation gene.

### Discovery of RNA editing sites

Previous studies identified 127 sites in 55 *Drosophila* genes that undergo A-to-I RNA editing<sup>41</sup>. This post-transcriptional modification is catalysed by dADAR, which is expressed at increasing levels throughout development and is thought to target products involved in nervous system function. We analysed the poly(A)<sup>+</sup> RNA-Seq data to identify exonic nucleotide positions consistent with A-to-I editing and defined 972 edited positions within transcripts of 597 genes, including previously described edited sites in the transcripts of 36

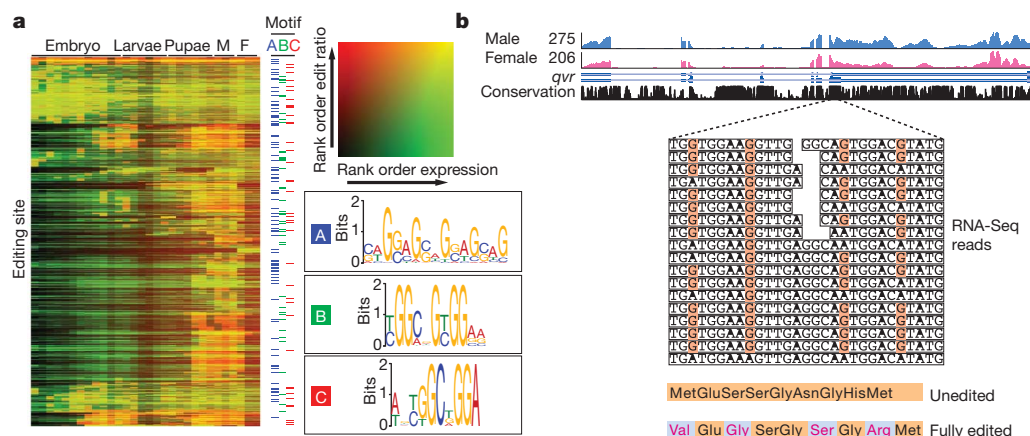
genes (Supplementary Table 29). These genes include those required for rapid neurotransmission and other widely ranging functions. For most sites, the frequency of editing increases throughout development and does not correlate with overall expression levels (Fig. 5a). Editing typically begins in late pupal stages, although we find transcripts that seem to be edited in late embryogenesis. Consistent with earlier studies<sup>42</sup>, exons containing editing sites are more highly conserved than unedited exons. The majority of the edited positions (630) alter amino acid coding, the others are either silent (201) or within untranslated regions (141). For example, the transcripts of *quiver* (*qvr*) are edited at six positions, four that result in amino acid changes (Fig. 5b). *qvr* encodes a potassium channel subunit that modulates the function of the voltage-gated Shaker (SH) potassium channel. *Sh* transcripts are also edited at multiple positions<sup>43</sup>. The combinatorial editing of both proteins probably has an important role in modulating action potentials in the arthropod nervous system and may have implications for the regulation of sleep<sup>44</sup>. Expressed sequence tags, long poly(A)<sup>+</sup> RNA-Seq and cDNAs cross-validate nearly one-quarter (214) of the newly discovered sites.

Computational analysis identified three potential editing-associated sequence motifs (Fig. 5a). We observe 381 sites with one or more motifs in close proximity to the edited nucleotide (Supplementary Table 30). Motif C, although less common than motifs A and B, is more strongly associated with the editing site. Most (93%) instances of motif C occur on the sense strand of the transcript and the A at the 3' end of the motif is the edited nucleotide. This motif is over-represented in editing events that occur early in development.

### Discussion

Our interrogation of the transcriptome of *D. melanogaster* throughout development has considerably expanded the number of building blocks used to make a fly. Specifically, we identified nearly 2,000 NTRs, increased the number of alternative splicing events by threefold and the number of RNA editing sites by an order of magnitude. The resulting view of the transcriptome at single-base resolution markedly improves our understanding of expression dynamics throughout the *Drosophila* life cycle and has substantial biological implications.

The *D. melanogaster*, *C. elegans* and human genomes are organized quite differently. Specifically, 20%, 45% and 2.5% of the *D. melanogaster*, *C. elegans* and human genomes, respectively, encode exons or mature transcripts. Primary transcripts comprise a larger fraction of each genome—60%, 82% and 37%. This highlights the fact that primary transcripts and introns are much shorter in *D. melanogaster* and *C. elegans* than in human and that the *D. melanogaster* and *C. elegans* genomes are more compact than the human genome.



**Figure 5 | Discovery of RNA editing events.** **a**, Rows represent edited sites. Rank-ordered expression levels (number of reads) are shown in green and the rank-ordered editing ratios are shown in red. Pictogram representations of editing motifs A, B and C are shown. **b**, RNA editing of *qvr*. Male and female

expression and conservation tracks are shown above RNA-Seq reads from adult females that align to the edited positions (orange). Conceptual translation of the unedited and fully edited transcripts result in four amino acid changes (red) at the C terminus of QVR.

The existence of unannotated genes was indicated by microarray studies<sup>8,45</sup> and conservation among *Drosophilid* genomes<sup>25</sup>. However, the NTRs that we identified were not identified by comparative sequence analysis<sup>46</sup> as they are less conserved than most previously known genes. This emphasizes the importance of using both comparative analyses and transcriptome profiling for genome annotation.

Despite the depth of our sequencing, the annotation of the *D. melanogaster* transcriptome is not finished. We failed to detect expression of 1,488 annotated genes including members of gene families to which short reads can not be uniquely mapped and genes expressed at low levels or in spatially and temporally restricted patterns. Moreover, although we substantially increased the fraction of genes that encode alternatively spliced or edited transcripts, we again failed to detect several annotated RNA processing events. Study of more temporally and spatially restricted samples will allow deeper exploration of the *Drosophila* transcriptome, and almost certainly result in the discovery of yet additional features. Furthermore, functional studies of the new and previously unstudied elements will provide valuable insight into metazoan development.

## METHODS SUMMARY

**Animal staging, collection and RNA extraction.** Isogenic ( $y^1$ ;  $cn\ bw^1\ sp^1$ ) embryos were collected at 2-h intervals for 24 h. Collection of later staged animals started with synchronized embryos and included resynchronizing with appropriate age indicators. Six larval, six pupal and three adult sexed stages, 1, 5 and 30 days, were collected. RNA was isolated using TRIzol (Invitrogen), DNase and purified on an RNeasy column (Qiagen). poly(A)<sup>+</sup> RNA was prepared from an aliquot of each total RNA sample using an Oligotex kit (Qiagen).

**Tiling arrays.** RNAs from three biological replicates of each sample were independently hybridized on 38-bp arrays (Affymetrix GeneChip *Drosophila* Tiling 2.0R array) as described<sup>47</sup>.

**RNA-Seq.** Libraries were generated and sequenced on an Illumina Genome Analyser IIx using single or paired-end chemistry and 76-bp cycles. SOLiD sequencing used total RNA treated with the RiboMinus Eukaryote Kit (Invitrogen). Samples were fragmented, adaptors ligated (Ambion) and sequenced for 50 bases using SOLiD V3 chemistry. 454 sequencing used poly(A)<sup>+</sup> RNA from Oregon R adult males and females and mixed-staged  $y^1$ ;  $cn\ bw^1\ sp^1$  embryos. Sequences are available from the Short Read Archive and the modENCODE website (<http://www.modencode.org/>).

**Targeted RT-PCR and cDNA isolation and sequencing.** Standard procedures were used for RT-PCR and targeted cDNA isolation and sequencing.

**Analysis.** Cufflinks<sup>24</sup> was used to identify new transcript models and to calculate expression levels for annotated and predicted transcript models. MFold<sup>48</sup> was used to predict secondary structures from the new snoRNA-like RNAs. JuncBASE<sup>49</sup> identified alternative splicing events and calculated per cent spliced in ( $\Psi$ )<sup>38</sup>. Editing sites were identified by comparing aligned reads to the reference genome.

Received 3 September; accepted 2 December 2010.

Published online 22 December 2010.

- Morgan, T. H. Sex limited inheritance in *Drosophila*. *Science* **32**, 120–122 (1910).
- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Nüsslein-Volhard, C. & Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**, 795–801 (1980).
- Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
- Spradling, A. C. Learning the common language of genetics. *Genetics* **174**, 1–3 (2006).
- Arbeitman, M. N. *et al.* Gene expression during the life cycle of *Drosophila melanogaster*. *Science* **297**, 2270–2275 (2002).
- Stolc, V. *et al.* A gene expression map for the euchromatic genome of *Drosophila melanogaster*. *Science* **306**, 655–660 (2004).
- Manak, J. R. *et al.* Biological function of unannotated transcription during the early development of *Drosophila melanogaster*. *Nature Genet.* **38**, 1151–1158 (2006).
- Bass, B. L. *RNA Editing* (Oxford Univ. Press, 2001).
- Rana, T. M. Illuminating the silence: understanding the structure and function of small RNAs. *Nature Rev. Mol. Cell Biol.* **8**, 23–36 (2007).
- Lipshitz, H. D., Peattie, D. A. & Hogness, D. S. Novel transcripts from the *Ultrathorax* domain of the Bithorax Complex. *Genes Dev.* **1**, 307–322 (1987).
- Meller, V. H., Wu, K. H., Roman, G., Kuroda, M. I. & Davis, R. L. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell* **88**, 445–457 (1997).
- Celniker, S. E. *et al.* Unlocking the secrets of the genome. *Nature* **459**, 927–930 (2009).
- The modENCODE Consortium *et al.* Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* doi:10.1126/science.1198374 (in the press).
- Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* doi:10.1126/science.1196914 (in the press).
- Bejerano, G. *et al.* A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* **441**, 87–90 (2006).
- Xie, X., Kamal, M. & Lander, E. S. A family of conserved noncoding elements derived from an ancient transposable element. *Proc. Natl Acad. Sci. USA* **103**, 11659–11664 (2006).
- Karch, F. *et al.* The abdominal region of the Bithorax Complex. *Cell* **43**, 81–96 (1985).
- Celniker, S. E., Sharma, S., Keelan, D. & Lewis, E. B. The molecular genetics of the bithorax complex of *Drosophila* cis-regulation in the *Abdominal-B* domain. *EMBO J.* **9**, 4277–4286 (1990).
- Ho, M. C. *et al.* Functional evolution of cis-regulatory modules at a homeotic gene in *Drosophila*. *PLoS Genet.* **5**, e1000709 (2009).
- Sanchez-Herrero, E. & Akam, M. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development* **107**, 321–329 (1989).
- Bae, E., Calhoun, V. C., Levine, M., Lewis, E. B. & Drewell, R. A. Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. *Proc. Natl Acad. Sci. USA* **99**, 16847–16852 (2002).
- Bender, W. MicroRNAs in the *Drosophila* bithorax complex. *Genes Dev.* **22**, 14–19 (2008).
- Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).
- Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
- Padgett, R. A. & Shukla, G. C. A revised model for U4atac/U6atac snRNA base pairing. *RNA* **8**, 125–128 (2002).
- Tycowski, K. T., Shu, M. D., Kukoyi, A. & Steitz, J. A. A conserved WD40 protein binds the Cajal body localization signal of scaRNP particles. *Mol. Cell* **34**, 47–57 (2009).
- Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* doi:10.1101/gr.112466.110 (in the press).
- Parisi, M. *et al.* A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biol.* **5**, R40 (2004).
- Andres, A. J. & Cherbas, P. Tissue-specific ecdysone responses: regulation of the *Drosophila* genes Eip28/29 and Eip40 during larval development. *Development* **116**, 865–876 (1992).
- Andres, A. J., Fletcher, J. C., Karim, F. D. & Thummel, C. S. Molecular analysis of the initiation of insect metamorphosis: a comparative study of *Drosophila* ecdysteroid-regulated transcription. *Dev. Biol.* **160**, 388–404 (1993).
- Lockett, T. J. & Ashburner, M. Temporal and spatial utilization of the alcohol dehydrogenase gene promoters during the development of *Drosophila melanogaster*. *Dev. Biol.* **134**, 430–437 (1989).
- Warren, J. T. *et al.* Discrete pulses of molting hormone, 20-hydroxyecdysone, during late larval development of *Drosophila melanogaster*: correlations with changes in gene activity. *Dev. Dyn.* **235**, 315–326 (2006).
- Costantino, B. F. *et al.* A novel ecdysone receptor mediates steroid-regulated developmental events during the mid-third instar of *Drosophila*. *PLoS Genet.* **4**, e1000102 (2008).
- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- van Bakel, H., Nislow, C., Blencowe, B. J. & Hughes, T. R. Most “dark matter” transcripts are associated with known genes. *PLoS Biol.* **8**, e1000371 (2010).
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J. & Blencowe, B. J. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **40**, 1413–1415 (2008).
- Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
- Philipps, D. L., Park, J. W. & Graveley, B. R. A computational and experimental approach toward a priori identification of alternatively spliced exons. *RNA* **10**, 1838–1844 (2004).
- Sanchez, L. Sex-determining mechanisms in insects. *Int. J. Dev. Biol.* **52**, 837–856 (2008).
- Stapleton, M., Carlson, J. W. & Celniker, S. E. RNA editing in *Drosophila melanogaster*: New targets and functional consequences. *RNA* **12**, 1922–1932 (2006).
- Jepson, J. E. & Reenan, R. A. Genetic approaches to studying adenosine-to-inosine RNA editing. *Methods Enzymol.* **424**, 265–287 (2007).
- Hoopengardner, B., Bhalla, T., Staber, C. & Reenan, R. Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**, 832–836 (2003).
- Wang, J. W. & Wu, C. F. Modulation of the frequency response of Shaker potassium channels by the quiver peptide suggesting a novel extracellular interaction mechanism. *J. Neurogenet.* **24**, 67–74 (2010).
- Hild, M. *et al.* An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5**, R3 (2003).
- Stark, A. *et al.* Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219–232 (2007).
- Cherbas, L. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* doi:10.1101/gr.112961.110 (in the press).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).



49. Brooks, A. N. *et al.* Conservation of an RNA regulatory map between *Drosophila* and mammals. *Genome Res.* doi:10.1101/gr.108662.110 (in the press).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Trapnell and L. Pachter for discussions and assistance with Cufflinks, and E. Clough for comments and feedback. A.N.B. was partially supported by an NSF graduate fellowship. This work was funded by an award from the National Human Genome Research Institute modENCODE Project (U01 HB004271) to S.E.C. (Principal Investigator) and M.R.B., P.C., T.R.G., B.R.G. and N.P. (co-Principal Investigators) under Department of Energy contract no. DE-AC02-05CH11231, and by the National Institute of Diabetes and Digestive and Kidney Diseases Intramural Research Program (B.O.).

**Author Contributions** J.A., M.R.B., P.C., T.R.G., B.R.G., R.A.H., T.C.K., B.O., N.P. and S.E.C. designed the project. J.A., S.E.B., M.R.B., P.C., T.R.G., B.R.G., R.A.H., B.O. and S.E.C. managed the project. D.M. prepared biological samples. T.C.K. oversaw biological sample production. D.Z. and B.E. prepared RNA samples. J.A. oversaw RNA sample production. W.L. and A.W. analysed array data. P.K. managed array data production. L.Y. prepared Illumina RNA-Seq libraries. C.A.D., L.L., J.E.S., K.H.W. and L.Y. performed Illumina sequencing. J.M.L., B.R.G. and S.E.C. managed Illumina sequencing

production. M.B. and R.E.G. performed 454 sequencing of adults. R.A.H. managed production of the embryonic SOLiD and 454 sequencing. C.A.D. managed data transfers. C.Z. managed databases and formatted array and sequence data for submission. C.G.A., P.J.B., S.E.B., A.N.B., S.D., M.O.D., B.R.G. and D.S. developed analysis methods. C.G.A., J.B.B., N.B., B.W.B., S.E.B., A.N.B., J.W.C., S.E.C., L.C., P.C., C.A.D., A.D., M.O.D., B.R.G., R.L., J.H.M., N.R.M., D.S. and Yi.Z. analysed data. B.B.T. aligned the SOLiD data. M.J.V. and J.M.L. generated annotations. C.G.A., D.S. and J.H.M. analysed species validation data. L.J., C.G.A., D.S. and N.R.M. performed species RNA-Seq quality control. Yu.Z. and J.H.M. oversaw sequencing and gathered species samples. C.G.A., A.N.B., J.W.C., L.C., P.C., A.H., D.S., J.M.L., R.L., N.R.M., J.H.M. and B.O. contributed to the text. A.H. assisted with manuscript preparation. B.R.G. and S.E.C. wrote the paper with input from all authors. All authors discussed the results and commented on the manuscript.

**Author Information** All sequence data have been deposited in the SRA, cDNA sequences have been deposited in GenBank, and array data deposited in GEO (see Supplementary Table 35 for all accession numbers). All data is also available at <http://www.modencode.org>. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to B.R.G. ([graveley@neuron.uchc.edu](mailto:graveley@neuron.uchc.edu)) or S.E.C. ([celniker@fruitfly.org](mailto:celniker@fruitfly.org)).

# Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*

Peter V. Kharchenko<sup>1,2</sup>, Artyom A. Alekseyenko<sup>3,4</sup>, Yuri B. Schwartz<sup>5†</sup>, Aki Minoda<sup>6</sup>, Nicole C. Riddle<sup>7</sup>, Jason Ernst<sup>8,9</sup>, Peter J. Sabo<sup>10</sup>, Erica Larschan<sup>3,4,11</sup>, Andrey A. Gorchakov<sup>3,4</sup>, Tingting Gu<sup>7</sup>, Daniela Linder-Basso<sup>5†</sup>, Annette Plachetka<sup>3,4</sup>, Gregory Shanower<sup>5†</sup>, Michael Y. Tolstorukov<sup>1,2</sup>, Lovelace J. Luquette<sup>1</sup>, Ruibin Xi<sup>1</sup>, Youngsook L. Jung<sup>1,3</sup>, Richard W. Park<sup>1,12</sup>, Eric P. Bishop<sup>1,12</sup>, Theresa K. Canfield<sup>10</sup>, Richard Sandstrom<sup>10</sup>, Robert E. Thurman<sup>10</sup>, David M. MacAlpine<sup>13</sup>, John A. Stamatoyannopoulos<sup>10,14</sup>, Manolis Kellis<sup>8,9</sup>, Sarah C. R. Elgin<sup>7</sup>, Mitzi I. Kuroda<sup>3,4</sup>, Vincenzo Pirrotta<sup>5</sup>, Gary H. Karpen<sup>6\*</sup> & Peter J. Park<sup>1,2,3\*</sup>

**Chromatin is composed of DNA and a variety of modified histones and non-histone proteins, which have an impact on cell differentiation, gene regulation and other key cellular processes. Here we present a genome-wide chromatin landscape for *Drosophila melanogaster* based on eighteen histone modifications, summarized by nine prevalent combinatorial patterns. Integrative analysis with other data (non-histone chromatin proteins, DNase I hypersensitivity, GRO-Seq reads produced by engaged polymerase, short/long RNA products) reveals discrete characteristics of chromosomes, genes, regulatory elements and other functional domains. We find that active genes display distinct chromatin signatures that are correlated with disparate gene lengths, exon patterns, regulatory functions and genomic contexts. We also demonstrate a diversity of signatures among Polycomb targets that include a subset with paused polymerase. This systematic profiling and integrative analysis of chromatin signatures provides insights into how genomic elements are regulated, and will serve as a resource for future experimental investigations of genome structure and function.**

The model organism Encyclopedia of DNA Elements (modENCODE) project is generating a comprehensive map of chromatin components, transcription factors, transcripts, small RNAs and origins of replication in *Drosophila melanogaster* and *Caenorhabditis elegans*<sup>1,2</sup>. *Drosophila* has been used as a model system for over a century to study chromosome structure and function, gene regulation, development and evolution. The availability of high-quality euchromatic and heterochromatic sequence assemblies<sup>3–5</sup>, extensive annotation of functional elements<sup>6</sup>, and a vast repertoire of experimental manipulations enhance the value of epigenomic studies in *Drosophila*.

Genome-wide profiling of chromatin components provides a rich annotation of the potential functions of the underlying DNA sequences. Previous work has identified patterns of post-translational histone modifications and non-histone proteins associated with specific elements (for example, transcription start sites, enhancers), as well as delineating the transcriptional status of genes and large domains<sup>7,8</sup>. Here we present a comprehensive picture of the chromatin landscape in a model eukaryotic genome. We define combinatorial chromatin ‘states’ at different levels of organization, from individual regulatory units to the chromosome level, and relate individual states to genome functions.

## Combinatorial chromatin states

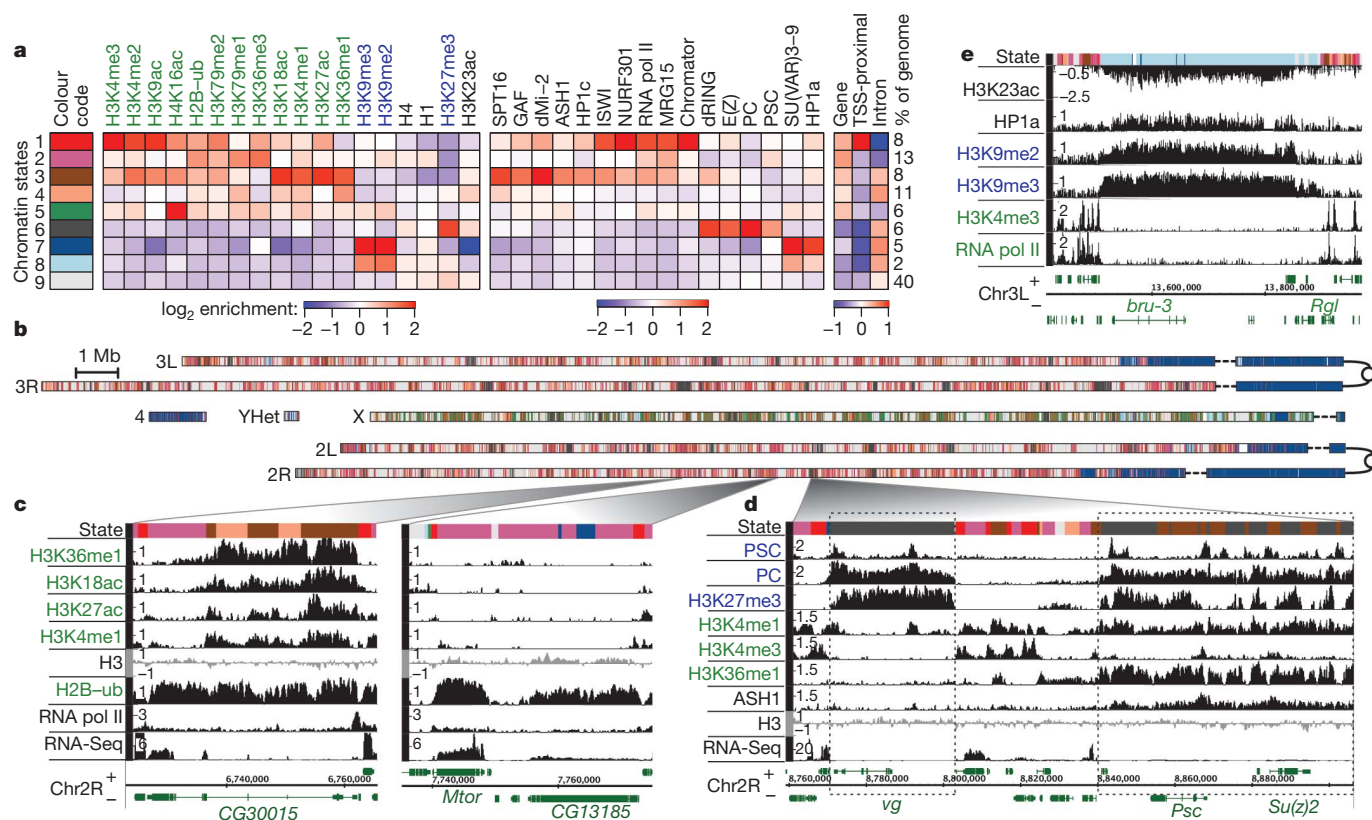
We performed chromatin immunoprecipitation (ChIP)-array analysis for numerous histone modifications and chromosomal proteins

(Supplementary Table 1), using antibodies tested for specificity and cross-reactivity<sup>9</sup> (Supplementary Fig. 1). Here we describe analyses of cell lines S2-DRSC (S2) and ML-DmBG3-c2 (BG3), derived from late male embryonic tissues (stages 16–17) and the central nervous system of male third instar larvae, respectively (see <http://www.modencode.org> for data from other cell lines and animal stages). Analysis reveals groups of correlated features, including those associated with heterochromatic regions<sup>10</sup>, Polycomb-mediated repression<sup>11</sup>, and active transcription<sup>12</sup> (Supplementary Fig. 2), similar to those observed in other organisms<sup>13,14</sup>. This indicates that specific histone modifications work together to achieve distinct chromatin ‘states’.

We used a machine-learning approach to identify the prevalent combinatorial patterns of 18 histone modifications, capturing the overall complexity of chromatin profiles observed in S2 and BG3 cells with 9 combinatorial states (Fig. 1a, Methods). The model associates each genomic location with a particular state, generating a chromatin-centric annotation of the genome (Fig. 1b). We examined each state for enrichment in non-histone proteins (Fig. 1a and Supplementary Fig. 3) and gene elements, as well as distribution across the karyotype (Fig. 1b and Supplementary Fig. 4) and finer-scale levels (Fig. 1c–e).

Most distinct chromatin states are associated with transcriptionally active genes. Active promoter and transcription start site (TSS)-proximal regions are identified by state 1 (Fig. 1; red), marked by prominent enrichment in H3K4me3/me2 (tri/dimethylation of residue K4 of

<sup>1</sup>Center for Biomedical Informatics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>2</sup>Children’s Hospital Informatics Program, Boston, Massachusetts 02115, USA. <sup>3</sup>Division of Genetics, Department of Medicine, Brigham & Women’s Hospital, Boston, Massachusetts 02115, USA. <sup>4</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>5</sup>Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854, USA. <sup>6</sup>Department of Molecular and Cell Biology, University of California at Berkeley, and Department of Genome Dynamics, Lawrence Berkeley National Lab, Berkeley, California 94720, USA. <sup>7</sup>Department of Biology, Washington University in St Louis, St Louis, Missouri 63130, USA. <sup>8</sup>MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139, USA. <sup>9</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>10</sup>Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>11</sup>Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island 02906, USA. <sup>12</sup>Graduate Program in Bioinformatics, Boston University, Boston, Massachusetts 02115, USA. <sup>13</sup>Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>14</sup>Department of Medicine, University of Washington, Seattle, Washington 98195, USA. †Present addresses: Department of Molecular Biology, Umea University, 901 87 Umea, Sweden. (Y.B.S.); Department of Plant Biology and Pathology, SEBS, Rutgers University, New Brunswick, New Jersey 08901, USA (D.L.-B.); Department of Basic Sciences, The Commonwealth Medical College, Scranton, Pennsylvania 18510, USA (G.S.). \*These authors contributed equally to this work.



**Figure 1 | Chromatin annotation of the *Drosophila melanogaster* genome.**

**a**, A 9-state model of prevalent chromatin states found in S2 and BG3 cells. Each chromatin state (row) is defined by a combinatorial pattern of enrichment (red) or depletion (blue) for specific chromatin marks (first panel, columns; active marks in green, repressive in blue). For instance, state 1 is distinguished by enrichment in H3K4me2/me3 and H3K9ac, typical of transcription start sites (TSS) in expressed genes. The enrichments/depletions are shown relative to chromatin input (S2 data shown, see Supplementary Fig. 3 for BG3 data and histone density normalization). The second panel shows average enrichment of chromosomal proteins. The third panel shows fold over/under-representation of genic and TSS-proximal ( $\pm 1$  kb) regions relative to the entire tiled genome. The enrichment of intronic regions is relative to genic regions associated with each state. **b**, A genome-wide karyotype view of the domains defined by the 9-state model in S2 cells. Centromeres are shown as open circles, and dashed

lines span gaps in the genome assembly. Several prominent chromatin organization features are illustrated (colour code in **a**), including the extent of pericentromeric heterochromatin (state 7) and the H4K16ac-driven signature of the dosage-compensated male X chromosome (state 5). (**BG3** in Supplementary Fig. 4.) **c–e**, Examples of chromatin annotation at specific loci. **c**, Two distinct chromatin signatures of transcriptionally active genes: one (left) is associated with enrichment in marks of states 3 and 4, whereas the other (right) is limited to states 1 and 2, recapitulating well established TSS and elongation signatures (note that small patches of state 7 in CG13185 illustrate H3K9me2 found at some expressed genes in S2 cells<sup>16</sup>). **d**, A locus containing two Polycomb-associated domains, silent (left) and balanced (right). **e**, A large state 8 domain located within euchromatic sequence in BG3 cells, enriched for chromatin marks typically associated with heterochromatic regions, but at lower levels than in pericentromeric heterochromatin (state 7).

histone H3) and H3K9ac (acetylation of K9 of histone H3). The transcriptional elongation signature associated with H3K36me3 enrichment is captured by state 2 (purple), found preferentially over exonic regions of transcribed genes. State 3 (brown), typically found within intronic regions, is distinguished by high enrichment in H3K27ac, H3K4me1 and H3K18ac. A related chromatin signature is captured by state 4 (coral), distinguished by enrichment of H3K36me1, but notably lacking H3K27ac. The number of genes associated with each chromatin state and the distribution of states within genes are shown in Supplementary Fig. 5.

Several aspects of large-scale organization are revealed by the karyotype view (Fig. 1b). Chromosome X is markedly enriched for state 5 (green), distinguished by high levels of H4K16ac in combination with H3K36me3 and other marks of 'elongation' state 2 (a combinatorial pattern associated with dosage compensation in male cells<sup>15</sup>). Pericentromeric heterochromatin domains and chromosome 4 are characterized by high levels of H3K9me2/me3 (state 7, dark blue)<sup>10</sup>. Finally, the model distinguishes another set of heterochromatin-like regions containing moderate levels of H3K9me2/me3 (state 8, light blue; Fig. 1e). Surprisingly, this state occupies extensive domains in autosomal euchromatic arms in BG3 cells, and in chromosome X in both cell lines<sup>16</sup>.

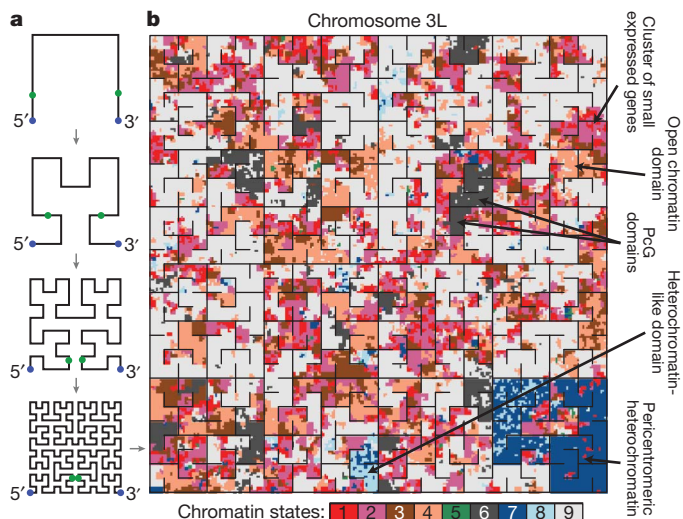
Further aspects of chromatin organization can be visualized by folding the chromosome using a Hilbert curve (Fig. 2a)<sup>17</sup>, which maintains the

spatial proximity of nearby elements. Thus, local patches of corresponding colours reveal the sizes and relative positions of domains associated with particular chromatin states (Fig. 2b and Supplementary Figs 6–9). For instance, specks of TSS-proximal regions (state 1) are typically contained within larger blocks of transcriptional elongation marks (state 2), which are in turn encompassed by extensive patches of H3K36me1-enriched domains (state 4) and variable-sized blocks of state 3. The clusters of open chromatin formed by these gene-centric patterns are separated by extensive silent domains (state 9) and regions of Polycomb-mediated repression (state 6). Factors responsible for domain boundaries were not identified in our analysis (Supplementary Fig. 10).

We also developed a multi-scale method to characterize chromatin organization at the spatial scale appropriate for the genome properties being investigated. For example, we observe that chromatin patterns most accurately reflect the replication timing of the S2 genome at scales of  $\sim 170$  kb (Supplementary Information, section 1). This is consistent with size estimates of chromatin domains influencing replication timing<sup>18</sup>, and suggests that multiple replication origins are coordinately regulated by the local chromatin environment (each replicon is  $\sim 28$ – $50$  kb<sup>19</sup>).

To examine combinatorial patterns not distinguished by the simplified 9-state model, we also generated a 30-state combinatorial model that uses presence/absence probabilities of individual marks<sup>20</sup> (Supplementary





**Figure 2 | Visualization of spatial scales and organization using compact folding.** **a**, The chromosome is folded using a geometric pattern (Hilbert space-filling curve) that maintains spatial proximity of nearby regions. An illustration of the first four folding steps is shown. Note that although this compact curve is optimal for preserving proximity relationships, some distal sites appear adjacent along the fold axis (green dots). **b**, Chromosome 3L in S2 cells. A domain of a given chromatin state appears as a patch of uniform colour of corresponding size. Thin black lines are used to separate regions that are distant on the chromosome. The folded view illustrates chromatin organization features that are not easily discerned from a linear view: active TSSs (state 1) appear as small specks surrounded by elongation state 2, commonly next to larger regions marked by H3K36me1-driven state 4, which also contains patches of intron-associated state 3. These open chromatin regions are separated by extensive domains of state 9. See Supplementary Figs 6 and 7 for other chromosomes and BG3 data. The folded views can be browsed alongside the linear annotations and other relevant data online: <http://compbio.med.harvard.edu/flychromatin>.

Fig. 11). The increased number of states can identify finer variations that are biologically significant, for example, a signature corresponding to transcriptional elongation in heterochromatic regions<sup>16</sup>.

### Chromatin state variation among genes

Active genes generally display enrichments or depletions of individual marks at specific gene segments (Fig. 3a). When classified according to their chromatin signatures (Supplementary Fig. 12), active genes fall into subclasses correlated with expression magnitude (Supplementary Information, section 2), gene structure and genomic

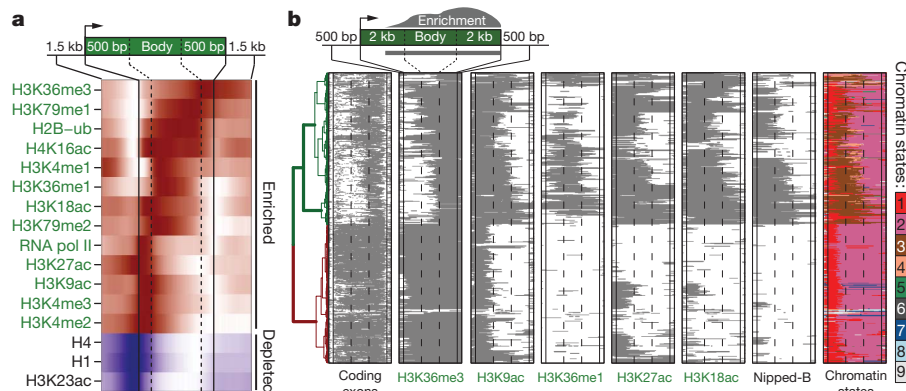
context (for example, heterochromatic genes combine H3K9me2/me3 with some active marks)<sup>16</sup>. Of particular interest is one class of long expressed genes, many with regulatory functions, which are enriched for H3K36me1 (cluster 2, Supplementary Fig. 12; 131 genes in S2, 202 in BG3; Supplementary Table 2).

To examine further the patterns associated with long genes, we clustered expressed autosomal genes  $\geq 4$  kb based on blocks of enrichment for each chromatin mark (Fig. 3b; 1,055 genes). We observe that genes with large 5'-end introns (green subtree, Fig. 3b; 552 genes) show extensive H3K27ac and H3K18ac enrichment, broader H3K9ac domains, and blocks of H3K36me1 enrichment (chromatin state 3, Fig. 3b, last column). These genes are enriched for developmental and regulatory functions (Supplementary Table 3), and are positioned within domains of Nipped-B<sup>21</sup> (Fig. 3b), a cohesin-complex loading protein previously associated with transcriptionally active regions<sup>21,22</sup>. In contrast, genes with more uniformly distributed coding regions (red subtree, Fig. 3b) lack most state 3 marks, and H3K9ac enrichment is restricted to the 2 kb downstream of the TSS. These differences are not explained by variation in histone density (Supplementary Fig. 13). Overall, the presence or absence of state 3 is the most common difference in the chromatin composition of expressed genes that are 1 kb and longer (Supplementary Fig. 14), and the presence of state 3 consistently correlates with a reduced fraction of coding sequence in the gene body, mainly associated with the presence of a long first intron.

State 3 domains are highly enriched for specific chromatin remodelling factors (SPT16 (also known as DRE4) and dMI-2; Supplementary Figs 15 and 16), whereas state 1 regions around active TSSs are preferentially bound by NURF301 (also called E(bx)) and MRG15. ISWI is enriched in both states 1 and 3 (Supplementary Figs 16 and 17). State 3 domains also exhibit the highest levels of nucleosome turnover<sup>23</sup>, and show higher enrichment of the transcription-associated H3.3 histone variant<sup>24</sup> than either the TSS- or elongation-associated states 1 and 2 (Supplementary Figs 15 and 16). Consistent with earlier analyses of cohesin-bound regions<sup>25</sup>, state 3 sequences tend to replicate early in G1 phase, and show abundance of early replicating origins (Supplementary Fig. 18). A regulatory role for state 3 domains is suggested by enrichment for a known enhancer binding protein (dCBP/p300<sup>26</sup>) in adult flies, and for enhancers validated in transgene constructs<sup>27</sup> (Supplementary Fig. 19).

### Modes of regulation in Polycomb domains

In *Drosophila*, loci repressed by Polycomb group (PcG) proteins are embedded in broad H3K27me3 domains that are regulated by Polycomb response elements (PREs) bound by E(Z), PSC and dRING (Fig. 1d)<sup>28,29</sup>. We find that regions of H3K4me1 enrichment surround all



**Figure 3 | Chromatin patterns associated with transcriptionally active genes.** **a**, Location and extent of chromatin features relative to boundaries of expressed genes ( $\geq 1$  kb) in BG3 cells. The colour intensity indicates the relative frequency of enrichment/depletion (red/blue) of a given mark within the gene (normalized independently for each mark). **b**, Regions enriched for 'active' chromatin marks in long transcribed genes. The plot shows the extent of regions enriched for various active marks at transcriptionally active genes ( $\geq 4$  kb) on

BG3 autosomes. Each row represents a scaled gene. The first column illustrates coding exons; the last column shows chromatin state annotation. The clustering of the genes according to the spatial patterns of chromatin marks separates genes with a high fraction of coding sequence (red subtree, bottom) from genes containing long introns (green subtrees, top), which are associated with chromatin state 3 (last column) and binding of specific chromosomal proteins, such as Nipped-B<sup>21</sup> (also see Supplementary Fig. 13).

PREs, 90% of which also display narrower peaks of H3K4me2 enrichment (Supplementary Fig. 20). Although this pattern is reminiscent of transcriptionally active promoter regions, PREs lack H3K4me3, indicating that a different mechanism of H3K4 methylation is used, perhaps involving the *Trithorax* H3K4 histone methyltransferase (HMTase) found at all PREs<sup>29</sup>.

To examine chromatin states associated with PcG targets, we analysed the chromatin and transcriptional signatures of TSSs in Polycomb-bound domains (Fig. 4a and Supplementary Fig. 21). In addition to fully repressed TSSs (cluster 1, Fig. 4a), we identify TSSs maintained in the 'balanced' state<sup>29</sup> (cluster 2, Fig. 4a), distinguished by coexistence of Polycomb with active marks (including the HMTase ASH1) and production of full-length messenger RNA transcripts (for example, *Psc* domain, Fig. 1d).

TSSs in clusters 3 and 4 are distinguished by the presence of adjacent PREs (Fig. 4a). Surprisingly, 53% of the PRE-proximal TSSs produce short RNA transcripts<sup>30</sup> (cluster 3, Fig. 4a), indicating stalling of engaged RNA pol II<sup>30</sup>. Using the global run-on sequencing (GRO-Seq) assay to accurately assess engaged RNA polymerases<sup>31</sup>, we observe that cluster 3 TSSs produce short transcripts in the sense orientation. The level of GRO<sup>+</sup> signal is similar to that found at fully transcribed genes (Supplementary Fig. 22); thus, some transcription initiates in cluster 3, but elongation fails. Interestingly, these genes are enriched for regulatory and developmental functions, even more than other genes within Polycomb domains (see Supplementary Tables 4 and 5). Genes without TSS-proximal PREs generally lack short transcript signatures (for example, cluster 1 in Fig. 4a; see Supplementary Fig. 21 for exceptions). Importantly, engaged polymerases and transcripts are not a general feature of PREs; TSS-distal PREs typically lack short RNA and GRO-Seq signals (Fig. 4b and Supplementary Fig. 22) despite being similarly enriched in H3K4me1/me2. The striking link between TSS-proximal PREs and the production of short RNAs suggests a potential mechanism for control of these developmental regulatory genes, whereby the same features that recruit H3K4 methyl marks to PREs may also facilitate RNA pol II recruitment to nearby TSSs.

## DHS plasticity and chromatin states

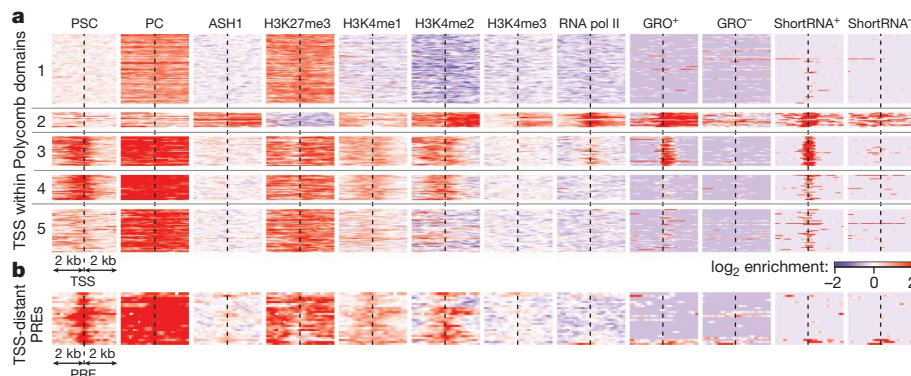
We used a DNase I hypersensitivity assay<sup>32,33</sup> to examine the distributions of putative regulatory regions and their relationships with chromatin states. DNase I hypersensitivity mapping broadly identifies sites with low nucleosome density and regions bound by non-histone proteins<sup>34–36</sup>. Short-read sequencing identified 8,616 high-magnitude DNase I hypersensitive sites (DHSs) in S2 cells and 6,354 in BG3 cells (and a comparable number of low-magnitude DHSs; Supplementary Fig. 23 and Methods). Approximately half of the high-magnitude DHSs are found at transcriptionally active TSSs (Supplementary Fig. 24). Thus, the chromatin context of the TSS-proximal DHSs is dominated by the features

expected for an active TSS, including RNA pol II, H3K4me3 and other state 1 marks (clusters 1, 2; Fig. 5a and Supplementary Fig. 25).

Of the 36% TSS-distal DHSs, most (60%) are positioned within annotated expressed genes (Supplementary Fig. 24). These gene-body DHSs are distinguished from TSS-proximal DHSs by low H3K4me3, higher levels of H3K4me1, H3K27ac, and other marks linked to chromatin state 3 (clusters 3, 4; Fig. 5a and Supplementary Fig. 26). An additional 20% of the TSS-distal DHSs are outside of annotated genes, but show signatures associated with active transcription starts or elongation, suggesting new alternative promoters or unannotated genes (Supplementary Figs 27 and 28). The remaining 20% of TSS-distal DHSs that appear to be intergenic (6% of all DHSs) are typically enriched for H3K4me1, but lack other active marks (cluster 5, Fig. 5a).

Most DHS positions fall into the TSS-proximal state 1 or the intron-biased state 3 (Fig. 5b). State 3 lacks H3K4me3 and is enriched for H3K4me1, H3K27ac and H3K18ac, similar to mammalian enhancer elements<sup>37</sup>. Many state 3 DHS positions are occupied by known regulatory proteins: GAGA factor binds to 49% of these DHSs in S2 cells, and developmental transcription factors bind to 44% of these DHSs in embryos<sup>38</sup>. Notably, we find that TSS-distal DHSs in *Drosophila* exhibit low-level bi-directional transcripts (Fig. 5a, shortRNA panel; see also Supplementary Figs 29 and 30), analogous to the enhancer RNAs (eRNAs) characterized in mice<sup>39</sup>. Analysis of GRO-Seq data (Fig. 5e) indicates that eRNA-like transcripts are common to both intra- and intergenic TSS-distal DHSs in *Drosophila*, a feature that is conserved with mammals.

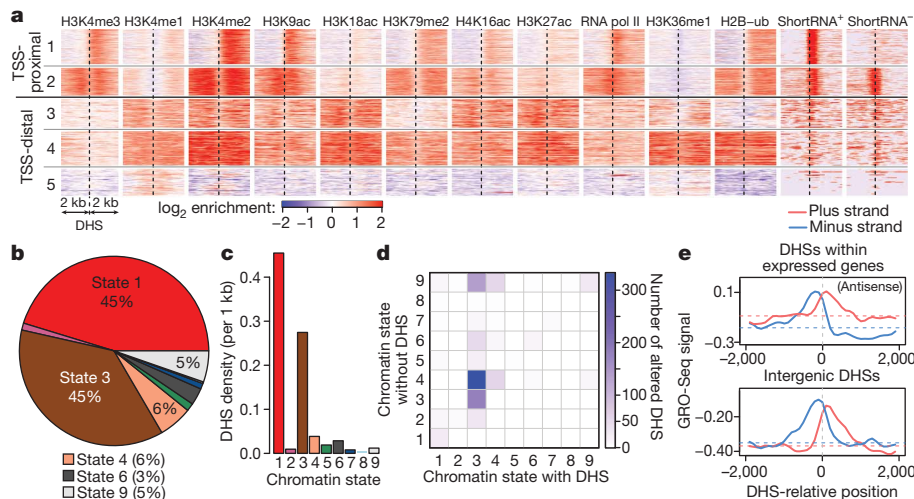
The association of DHSs with chromatin states 1 and 3 (Fig. 5c) persists even in chromosome 4 and pericentromeric heterochromatin, where such states are infrequent (Supplementary Fig. 31). This suggests that these chromatin states and associated remodelling factors (for example, ISWI, SPT16) provide the context necessary for non-histone chromosomal protein binding at DHSs, or are the consequence of such binding events. To investigate this interdependency, we analysed a high-confidence set of loci that exhibit DHSs in only one of the two examined cell lines (Supplementary Fig. 32). Surprisingly, although in general more DHSs are in state 1 regions, 91% of the cell-type-specific DHSs are found within state 3 domains (14-fold increase compared to state 1 DHSs; Supplementary Table 6 and Fig. 5d). Comparison with DHSs in an additional cell type (Kc167, Supplementary Fig. 33) confirms that DHSs displaying plasticity between cell types are mostly found in state 3. When DHSs are absent, the altered loci maintain chromatin state 3 in 23% of the cases (Fig. 5d), indicating that the presence of state 3 is not always dependent on the DHS. More frequently, the altered loci transition to state 4 (43% of the cases), an open chromatin state that lacks many of the histone modifications and chromatin remodellers characteristic of state 3. Although the less frequent



**Figure 4 | Signatures of TSSs within domains of Polycomb-mediated repression.** **a**, Distinct classes of TSSs in S2 cell Polycomb domains. Each row represents a TSS. Clusters 1–5 illustrate distinct TSS states (see Supplementary Fig. 21 for complete set of clusters). Cluster 1 shows fully repressed TSSs with the expected pattern of PC and H3K27me3 enrichment; cluster 2 shows 21 TSSs found within ASH1 domains, maintained in a balanced state. Clusters 3 and 4 distinguish TSSs located in the immediate proximity of Polycomb response

elements (PREs), showing the symmetrical H3K4me1/me2 enrichment typical of all PREs. Many such TSSs (cluster 3, 42 TSSs) produce short, non-polyadenylated transcripts along the sense strand (GRO<sup>+</sup>/shortRNA<sup>+</sup> columns), indicating the presence of paused polymerase. **b**, PRE positions distant from annotated TSSs. TSS-distal PREs exhibit enrichment for H3K4me1/me2, but are not associated with GRO or shortRNA signatures.





**Figure 5 | Chromatin signatures of regulatory elements identified by DNase I hypersensitivity.** **a**, Representative classes of high-magnitude DNase I hypersensitive sites (DHSs) and chromatin signatures in S2 cells. TSS-proximal (within 2 kb) DHSs show chromatin signatures expected of expressed gene promoters: high H3K4me3 and RNA pol II signal extending in the direction of transcription (left to right; cluster 2 groups bi-directional promoters). TSS-distal DHSs are associated with high H3K4me1 and low H3K4me3 levels. Most TSS-distal DHSs found within the bodies of expressed genes (clusters 3, 4) are associated with chromatin state 3. A cluster of rare intergenic DHSs (cluster 5) is associated with localized peaks of H3K4me1/2 (complete sets of clusters in Supplementary Figs 25, 26 and 28). **b**, Distribution of DHS positions among chromatin states. The vast majority of DHSs are found within the TSS-proximal state 1 or enhancer-like state 3 regions. **c**, States 1 and 3 exhibit the highest

density of DHSs. **d**, Cell-line-specific DHSs are positioned predominantly within the enhancer-like state 3. The transition matrix shows the chromatin state of loci containing DHSs in one cell line (*x* axis), and the state of the same locus in the other cell line where the DHS is absent (*y* axis). Most of the DHSs that differ between cell lines originate from state 3. When DHSs are absent, the loci typically transition to an open chromatin state 4 (43%), or maintain state 3 (23%). In both scenarios, most of the associated genes remain transcriptionally active (see Supplementary Fig. 34). **e**, Low levels of engaged RNA polymerase are associated with TSS-distal DHSs. The top plot shows the local increase in the antisense GRO-seq signal for DHSs located within transcribed genes; dashed lines show median levels. Intergenic DHS positions (bottom plot) also show bi-directional GRO-seq signal of comparable magnitude. See Supplementary Figs 27, 29 and 30.

transitions to the Polycomb state 6 (7%) or background state 9 (17%) typically coincide with gene silencing, most of the genes that maintain state 3 or transition to state 4 remain transcriptionally active (Supplementary Fig. 34). These observations provide further support for an enhancer-like function for state 3 DHSs, and suggest a more subtle regulatory role than simple linkage to the presence or absence of gene expression.

### Chromatin annotation of genome functions

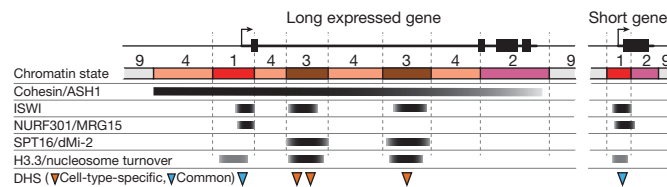
The genomic chromatin state annotation and discovery of refined chromatin signatures for chromosomes, domains, and subsets of regulatory genes demonstrate the utility of a systematic, genome-wide profiling of an organism that is already understood in considerable detail. Clearly, the definition and functional annotation of chromatin patterns will be enhanced by incorporation of data for different types of components. Five 'colours' of chromatin were recently identified in Kc167 cells using chromosomal protein maps<sup>40</sup>. Comparison with our 9-state model shows similarities as well as differences in the ability to distinguish functional elements (Supplementary Fig. 35); thus, further integration of such data in the same cell type may resolve additional functional features. Our results illustrate the utility of integrating multiple data types (histone marks, non-histone proteins, chromatin accessibility, short RNAs and transcriptional activity) for comprehensive characterization of functional chromatin states.

An important, repeated theme is that chromatin state analysis identifies unexpected distinctions between subsets of active genes. Besides the differences linked to genomic context (for example, male X chromosome, heterochromatin), the main source of variability is the presence of the acetylation-rich state 3 (Fig. 6). Several lines of evidence suggest that the intronic positions marked by state 3 are important for gene regulation. State 3 regions show specific associations with known chromatin remodellers (SPT16, dMi-2 and ISWI) and gene regulatory proteins (for example, GAF, dCBP/p300), and the highest rates of nucleosome turnover and transcription-dependent deposition of the H3.3 variant. State 3 genes are also bound by cohesin

complex proteins, thought to associate with decondensed chromatin<sup>21</sup> to promote looping interactions with promoter regions<sup>22</sup>.

A regulatory role for state 3 chromatin is further suggested by the high density of DHSs, comparable to that of active TSS state 1, and the fact that state 3 accounts for most of the DHS plasticity among cell types. The combinations of histone marks found in state 3 are similar to signatures of mammalian enhancers<sup>37</sup>, which also show high variability between cell types<sup>37</sup>. Furthermore, state 3 DHSs exhibit low levels of short, non-coding bidirectional transcripts reminiscent of eRNAs identified in mice<sup>39</sup>. Together, these findings suggest that state 3 regions contain enhancers or other regulatory elements, and that a combination of modifications can be used to identify new elements in the genome.

Genes within repressive Polycomb domains also display several distinct combinatorial chromatin patterns (Fig. 4a), which probably represent a range of functional states: repressed, paused, or expressed genes in either balanced<sup>29</sup> or fully activated states. Alternatively, distinct signatures might mark subsets of regulatory genes that require either long-term repression or the ability to reverse functional states, depending on environmental or developmental cues. The PRE-proximal paused TSSs have some similarity to the 'bivalent' genes in mammalian cells, which also display transcriptional pausing of key regulatory and developmental genes<sup>41,42</sup>. However, the mammalian 'bivalent state' is characterized by the simultaneous



**Figure 6 | Spatial arrangements of chromatin states associated with active transcription.** Unlike short or exon-rich expressed genes, expressed genes with long intronic regions commonly contain one or more regions of enhancer-like state 3, associated with specific chromosomal proteins, high nucleosome turnover and DHSs displaying cell-type plasticity.



presence of PcG proteins, H3K27me3 and H3K4me3, which in *Drosophila* is found only in the fully elongating 'balanced' state<sup>29,43</sup>.

Comprehensive analysis of chromatin signatures has enormous potential for annotating functional elements in both well studied and new genomes. Going forward, our systematic characterization of the epigenomic and transcriptional properties of *Drosophila* cells should spur in-depth experimental analyses of the relationship between chromatin states and genome functions, ranging from whole chromosomes down to individual regulatory elements and circuits.

## METHODS SUMMARY

Histone modification and chromosomal protein antibodies were characterized for cross-reactivity. ChIP-chip was performed in duplicate, using Affymetrix *Drosophila* Tiling 2.0R Arrays. Digital DNase I-Seq assays were performed as described previously<sup>44</sup>, and Global Run-On library (GRO-Seq) data was generated as described previously<sup>31</sup>. Short RNA data was generated by ref. 30, and RNA-Seq data was generated by ref. 45. See ref. 46 for other modENCODE RNA-Seq data. The chromatin state models were generated as hidden Markov models (HMMs) of different histone marks. DHSs were identified as read density peaks significantly enriched relative to the genomic DNA control. Clustering of chromatin signatures was determined using the PAM algorithm.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 2 September; accepted 6 December 2010.**

**Published online 22 December 2010; corrected 24 March 2011 (see full-text HTML version for details).**

- modENCODE. Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* doi:10.1126/science.1198374 (in the press).
- Gerstein, M. B. et al. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* doi:10.1126/science.1196914 (in the press).
- Adams, M. D. et al. The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
- Clark, A. G. et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).
- Hoskins, R. A. et al. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science* **316**, 1625–1628 (2007).
- Tweedie, S. et al. FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res.* **37**, D555–D559 (2009).
- Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
- Mendenhall, E. M. & Bernstein, B. E. Chromatin state maps: new technologies, new insights. *Curr. Opin. Genet. Dev.* **18**, 109–115 (2008).
- Egelhofer, T. A. et al. An assessment of histone-modification antibody quality. *Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1972 (5 December 2010).
- Eisenberg, J. C. & Reuter, G. Cellular mechanism for targeting heterochromatin formation in *Drosophila*. *Int. Rev. Cell Mol. Biol.* **273**, 1–47 (2009).
- Schwartz, Y. B. & Pirota, V. Polycomb complexes and epigenetic states. *Curr. Opin. Cell Biol.* **20**, 266–273 (2008).
- Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* **128**, 707–719 (2007).
- Liu, C. L. et al. Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol.* **3**, e328 (2005).
- Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Larschan, E. et al. MSL complex is attracted to genes marked by H3K36 trimethylation using a sequence-independent mechanism. *Mol. Cell* **28**, 121–133 (2007).
- Riddle, N. C. et al. Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res.* doi:10.1101/gr.110098.110 (in the press).
- Anders, S. Visualization of genomic data with the Hilbert curve. *Bioinformatics* **25**, 1231–1235 (2009).
- MacAlpine, D. M., Rodriguez, H. K. & Bell, S. P. Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* **18**, 3094–3105 (2004).
- Blumenthal, A. B., Kriegstein, H. J. & Hogness, D. S. The units of DNA replication in *Drosophila melanogaster* chromosomes. *Cold Spring Harb. Symp. Quant. Biol.* **38**, 205–223 (1974).
- Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnol.* **28**, 817–825 (2010).
- Misulovin, Z. et al. Association of cohesin and Nipped-B with transcriptionally active regions of the *Drosophila melanogaster* genome. *Chromosoma* **117**, 89–102 (2008).
- Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
- Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19**, 460–469 (2009).
- MacAlpine, H. K., Gordan, R., Powell, S. K., Hartemink, A. J. & MacAlpine, D. M. *Drosophila* ORC localizes to open chromatin and marks sites of cohesin complex loading. *Genome Res.* **20**, 201–211 (2010).
- Visel, A. et al. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* **462**, 65–70 (2009).
- Schwartz, Y. B. et al. Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nature Genet.* **38**, 700–705 (2006).
- Schwartz, Y. B. et al. Alternative epigenetic chromatin states of Polycomb target genes. *PLoS Genet.* **6**, e1000805 (2010).
- Nechaev, S. et al. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in *Drosophila*. *Science* **327**, 335–338 (2010).
- Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848 (2008).
- Wu, C. The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNase I. *Nature* **286**, 854–860 (1980).
- Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell* **16**, 797–806 (1979).
- Elgin, S. C. The formation and function of DNase I hypersensitive sites in the process of gene activation. *J. Biol. Chem.* **263**, 19259–19262 (1988).
- Jin, C. et al. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nature Genet.* **41**, 941–945 (2009).
- Hesselberth, J. R. et al. Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nature Methods* **6**, 283–289 (2009).
- Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- MacArthur, S. et al. Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10**, R80 (2009).
- Kim, T. K. et al. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Filion, G. J. et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
- Bernstein, B. E. et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Kanhere, A. et al. Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
- Schuettengruber, B. et al. Functional anatomy of polycomb and trithorax chromatin landscapes in *Drosophila* embryos. *PLoS Biol.* **7**, e13 (2009).
- Sekimata, M. et al. CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon- $\gamma$  locus. *Immunity* **31**, 551–564 (2009).
- Cherbas, L. et al. The transcriptional diversity of 25 *Drosophila* cell lines. *Genome Res.* **21**, doi:10.1101/gr.112961.110 (in the press).
- Gravely, B. R. et al. The developmental transcriptome of *Drosophila melanogaster*. *Nature* doi:10.1038/nature09715 (this issue).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank our technicians D. Acevedo, S. Gadel, C. Kennedy, O.-K. Lee, S. Marchetti, S. Vong and M. Weaver, and Rutgers BRTC. We also thank our colleagues who donated antibodies: J. Kadonaga (H1), A. L. Greenleaf (RNA pol II), G. Reuter (SU(VAR)3-9), G. Cavalli (GAF) and I. F. Zhimulev/H. Saumweber (Chromator). The major support for this work came from the modENCODE grant U01HG004258 to G.H.K. (Principal Investigator) and S.C.R.E., M.I.K., P.J.P. and V.P. (co-Principal Investigators), administered under Department of Energy contract no. DE-AC02-05CH11231. Additional funding came from RC2 HG005639, U01 HG004279, R01 GM082798, R37 GM45744, RC1 HG005334, R01 GM071923, U54 HG004592 and NSF 0905968.

**Author Contributions** P.V.K. performed most bioinformatic analysis. A.A.A., Y.B.S., A.M., N.C.R., E.L., A.A.G., T.G., D.L.-B., A.P. and G.S. generated data, directed by S.C.R.E., M.I.K., V.P. and G.H.K. The 30-state analysis was performed by J.E. and M.K., whereas M.Y.T., L.J.L., R.X., Y.L.J., R.W.P. and E.P.B. performed additional bioinformatic analysis. P.J.S., T.K.C., R.S., R.E.T. and J.A.S. generated and processed DHS data. D.M.M. helped with replication analysis. P.J.P. supervised all analysis. G.H.K. coordinated the entire project. P.V.K., G.H.K. and P.J.P. wrote the manuscript, with contributions from S.C.R.E., M.I.K., V.P., Y.B.S., N.C.R., A.A.A. and A.M.

**Author Information** The data are available from the modENCODE site (<http://www.modencode.org>). GRO-Seq data are available from Gene Expression Omnibus (GEO, GSE25321). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to P.J.P. ([peter\\_park@harvard.edu](mailto:peter_park@harvard.edu)) or G.H.K. ([karpem@fruitfly.org](mailto:karpem@fruitfly.org)).

## METHODS

**Growth conditions.** ML-DmBG3-c2 cells were obtained from DGRC (<https://dgrc.cgb.indiana.edu/>), and S2-DRSC cells were from the DRSC (<http://www.flyrnai.org/>). All cell lines were grown to a density of  $\sim 5 \times 10^6$  cells  $\text{ml}^{-1}$  in Schneider's media (Gibco) supplemented with 10% FCS (HyClone).  $10 \mu\text{g ml}^{-1}$  insulin was added to the ML-DmBG3-c2 media.

**Antibodies.** Antibodies are listed in Supplementary Table 1. Commercial antibodies against modified histones were tested by western blot for the lack of cross-reactivity with the corresponding recombinant histone produced in *Escherichia coli* and non-histone proteins from embryonic nuclear extracts. Antibody specificity was further assayed by western dot/slot blot against a panel of synthetic modified histone peptides. Only antibodies that showed  $<50\%$  of total signal associated with non-histone proteins, and more than fivefold higher affinity for the corresponding histone peptide, were used in ChIP experiments.

The specificity of antibodies against chromosomal proteins was tested by western blots with nuclear extracts prepared from mutant flies or S2 cells subjected to RNAi knockdown<sup>47</sup>. An antibody was considered specific if it recognized a major band of expected mobility that was absent in the sample prepared from mutant flies, or diminished twofold or more after RNAi depletion. When possible, distributions of a chromosomal protein were mapped with two antibodies generated against different epitopes (see Supplementary Fig. 17). Data from chromatin proteins for which only one antibody was available were validated by comparison with published genomic distributions for a different component of the same complex, or to published genomic distributions generated with a different antibody.

**ChIP and microarray hybridization.** Crosslinked chromatin from cultured cells was prepared as described<sup>28</sup> with the following modifications. Before ultrasound shearing, cells were permeabilized with 1% SDS, and shearing was done in TE-PMSF (0.1% SDS, 10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0, 1 mM PMSF) using a Bioruptor (Diagenode) ( $2 \times 10$  min,  $1 \times 5$  min; 30 s on, 30 s off; high power setting).

ChIP was performed as in ref. 28 and immunoprecipitated DNA was amplified using the whole genome amplification kit (WGA2, Sigma) according to the manufacturer's instructions (chemical fragmentation step was omitted). The amplified material was labelled and hybridized to *Drosophila* Tiling Arrays v2.0 (Affymetrix) as in ref. 28.

**Processing of ChIP data.** At least two independent biological replicates were assessed for each ChIP profile. The  $\log_2$  intensity ratios ( $M$  values) were calculated for each replicate. The profiles were smoothed using local regression (lowess) with 500 bp bandwidth, and the genome-wide mean was subtracted. The regions of significant enrichment were determined as clusters of at least 1 kb in length, with gaps no more than 100 bp where  $M$  value exceeds a statistically significant (0.1% false discovery rate (FDR)) enrichment threshold. The set of biological replicates was deemed consistent if the enriched regions from individual experiments had a 75% reciprocal overlap, or if at least 80% of the top 40% of the regions identified in each experiment were identified in the other replicate (before comparison the replicates were size-equalized by increasing the significance threshold for a replicate with more enriched sequence). The data from individual replicates were then combined using local regression smoothing, and used for all of the presented analysis, unless noted otherwise.

**DNase I hypersensitivity.** Digital DNase I-Seq assays were performed as described previously<sup>44</sup>. The sequenced reads were aligned to the Berkeley *Drosophila* Genome Project release 5 (BDGP.R5) genome assembly, recording only uniquely mappable reads. To detect DNase I hypersensitive sites, hotspot positions were identified based on a 300-bp scanning window statistic (Poisson model relative to 50 kb background density, Z-score threshold of 2), and peaks of read density were selected within the hotspots using randomization-based thresholding at 0.1% FDR. The set of high-magnitude DHSs analysed here (except for Supplementary Fig. 23) was identified as a subset of all peaks that show statistically significant enrichment over the normalized genomic DNA read density profile (using a 300-bp window

centred around the peak, binomial model, with Z-score threshold of 3). This method controls for copy number variation and sequencing/mapping biases; however, it may also reduce the sensitivity of DHS detection. In the DHS chromatin profile clustering analysis (Fig. 5a, relevant Supplementary figures), DHSs found within 1 kb of another DHS were excluded if their enrichment magnitude (relative to genomic background) was lower (to avoid showing the same region more than once).

**RNA sequencing.** The preparation of RNA-Seq libraries and sequencing is described in ref. 45. The sequenced reads were aligned to the BDGP.R5 genome assembly and annotated exon junctions, recording only uniquely mappable reads. The RPKM (reads per kilobase of exonic sequence per million reads mapped) was estimated for each exon. The total transcriptional output of each annotated gene was estimated based on the maximum of all exons within the gene. The presented analysis uses  $\log_{10}(\text{RPKM} + 1)$  values unless otherwise noted.

**GRO sequencing.** Global Run-On library was prepared from S2 cells and sequenced as described<sup>31</sup>. The reads were aligned to the BDGP.R5 genome assembly, recording only uniquely mappable reads. The smoothed profiles of reads mapping to each strand were calculated using Gaussian smoothing ( $\sigma = 100$  bp). The analysis uses  $\log_{10}(d + 1)$ , where  $d$  is the smoothed density value.

**Short RNA data processing.** The short RNA data for S2 cells was generated by ref. 30, and was aligned and processed in the same way as the GRO-Seq data.

**Chromatin state models.** To derive a 9-state joint chromatin state model for S2 and BG3 cells (Fig. 1a), the genome was first divided into 200-bp bins, and the average enrichment level was calculated within each bin based on unsmoothed  $\log_2$  intensity ratio values taking into account individual replicates, using all histone enrichment profiles and PC to discount the genome-wide difference in S2 H3K27me3 profiles. The bin-average values of each mark were shifted by the genome-wide mean, scaled by the genome-wide variance, and quantile-normalized between the two cells. The hidden Markov model (HMM) with multivariate normal emission distributions was then determined from the Baum-Welch algorithm using data from both cell types, and 30 seeding configurations determined with K-means clustering. States with minor intensity variations (Euclidian distance of mean emission values  $<0.15$ ) were merged. Larger models (up to 30 states) were examined, and the final number of states was chosen for optimal interpretability.

An extensive discrete chromatin state model (Supplementary Fig. 11) was calculated as described in ref. 20. The model was trained using a 200-bp grid with binary calls (enriched/not enriched). The binary calls were made based on a 5% FDR threshold determined from ten genome-wide randomizations for each mark. For H1, H4 and H3K23ac regions of significant depletion rather than enrichment were called.

**Regions of enrichment for individual marks.** To determine contiguous regions of enrichment for individual marks, a three-state HMM was used, with states corresponding to enriched, neutral and depleted profiles (normally-distributed emission parameters:  $(\mu = [-0.5 \ 0 \ 0.5], \sigma^2 = 0.3)$ ). The enriched regions were determined from the Viterbi path. The HMM segmentation was applied to unsmoothed  $M$  value data taking into account individual biological replicates. The genes were clustered based on the combinatorial pattern of occurrence of enriched regions (coding exons and state panels were not used for clustering).

**Classification of enrichment profiles.** Clustering of chromatin signatures around TSSs (Fig. 4a), PREs (Fig. 4b) and DHSs (Fig. 5a and relevant Supplementary Information sections) was determined using the Partitioning Around Medoids algorithm. For clustering, each profile was summarized with average values within bins spanning  $\pm 2$ -kb regions. One-hundred-base-pair bins were used for the central  $\pm 500$ -bp region, 300-bp bins outside.

47. Clemens, J. C. *et al.* Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA* **97**, 6499–6503 (2000).

# Interaction-based quantum metrology showing scaling beyond the Heisenberg limit

M. Napolitano<sup>1</sup>, M. Koschorreck<sup>1</sup>, B. Dubost<sup>1,2</sup>, N. Behbood<sup>1</sup>, R. J. Sewell<sup>1</sup> & M. W. Mitchell<sup>1</sup>

Quantum metrology aims to use entanglement and other quantum resources to improve precision measurement<sup>1</sup>. An interferometer using  $N$  independent particles to measure a parameter  $\mathcal{X}$  can achieve at best the standard quantum limit of sensitivity,  $\delta\mathcal{X} \propto N^{-1/2}$ . However, using  $N$  entangled particles and exotic states<sup>2</sup>, such an interferometer<sup>3</sup> can in principle achieve the Heisenberg limit,  $\delta\mathcal{X} \propto N^{-1}$ . Recent theoretical work<sup>4–6</sup> has argued that interactions among particles may be a valuable resource for quantum metrology, allowing scaling beyond the Heisenberg limit. Specifically, a  $k$ -particle interaction will produce sensitivity  $\delta\mathcal{X} \propto N^{-k}$  with appropriate entangled states and  $\delta\mathcal{X} \propto N^{-(k-1/2)}$  even without entanglement<sup>7</sup>. Here we demonstrate ‘super-Heisenberg’ scaling of  $\delta\mathcal{X} \propto N^{-3/2}$  in a nonlinear, non-destructive<sup>8,9</sup> measurement of the magnetization<sup>10,11</sup> of an atomic ensemble<sup>12</sup>. We use fast optical nonlinearities to generate a pairwise photon–photon interaction<sup>13</sup> (corresponding to  $k = 2$ ) while preserving quantum-noise-limited performance<sup>7,14</sup>. We observe super-Heisenberg scaling over two orders of magnitude in  $N$ , limited at large numbers by higher-order nonlinear effects, in good agreement with theory<sup>13</sup>. For a measurement of limited duration, super-Heisenberg scaling allows the nonlinear measurement to overtake in sensitivity a comparable linear measurement with the same number of photons. In other situations, however, higher-order nonlinearities prevent this crossover from occurring, reflecting the subtle relationship between scaling and sensitivity in nonlinear systems. Our work shows that interparticle interactions can improve sensitivity in a quantum-limited measurement, and experimentally demonstrates a new resource for quantum metrology.

The most precise instruments are interferometric in nature, and operate according to the laws of quantum mechanics. A collection of particles, for example photons or atoms, is prepared in a superposition state, allowed to evolve under the action of a Hamiltonian containing an unknown parameter,  $\mathcal{X}$ , and measured in agreement with quantum measurement theory. The complementarity of quantum measurements<sup>15</sup> determines the ultimate sensitivity of these instruments.

Here we describe polarization interferometry, used, for example, in optical magnetometry to detect atomic magnetization<sup>11,16,17</sup>; similar theory describes other interferometers<sup>3</sup>. A collection of  $N$  photons, with circular plus- and minus-polarization eigenstates,  $|+\rangle$  and  $|-\rangle$ , is described by single-photon Stokes operators  $\hat{S}_i = (1/2)(|+\rangle, |-\rangle)\sigma_i(|+\rangle, |-\rangle)^T$ , where  $\sigma_i$  ( $i = x, y, z$ ) are the Pauli matrices,  $\sigma_0$  is the identity and a superscript ‘T’ denotes transposition. In traditional quantum metrology, a Hamiltonian of the form  $\hat{H} = \hbar\mathcal{X} \sum_{j=1}^N \hat{S}_z^{(j)}$ , where  $\hbar$  denotes Planck’s constant divided by  $2\pi$ , uniformly and independently couples the photons to  $\mathcal{X}$ , the parameter to be measured<sup>1</sup>. If the input state consists of independent photons, the possible precision scales as  $\delta\mathcal{X} \propto N^{-1/2}$ , the shot noise or standard quantum limit (SQL). The factor of  $N^{-1/2}$  reflects the statistical averaging of independent results. In contrast, entangled states can be highly, even perfectly, correlated, giving precision limited by  $\delta\mathcal{X} \propto N^{-1}$ , the Heisenberg limit.

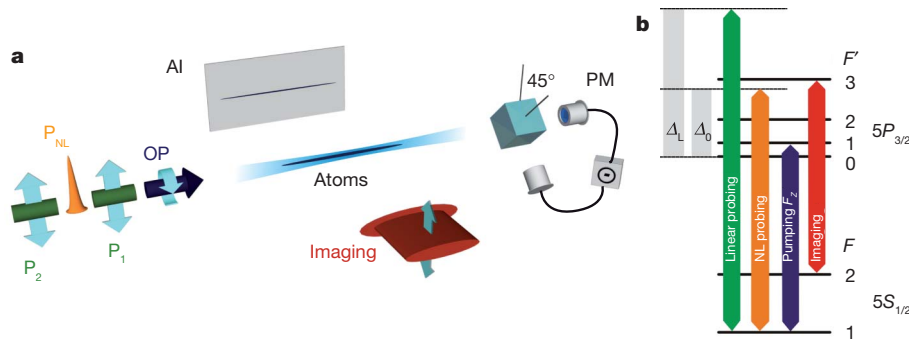
The above Hamiltonian is conveniently written  $\hat{H} = \hbar\mathcal{X}\hat{S}_z$ , where  $\hat{S}_i \equiv \sum_{j=1}^N \hat{S}_i^{(j)}$  is a collective variable describing the net polarization of the photons. The independence of the photons manifests itself in the linearity of this Hamiltonian. Recently, it has been shown that interactions among particles, or, equivalently, nonlinear Hamiltonians, can contribute to measurement sensitivity and give scaling beyond the Heisenberg limit<sup>4</sup>. For example, a Hamiltonian  $\hat{H} = \hbar\mathcal{X}\hat{S}_z^k$ , that is, with a  $k$ th-order nonlinearity in  $\hat{S} \equiv (\hat{S}_x, \hat{S}_y, \hat{S}_z)$ , contains  $k$ -photon interaction terms  $\hat{S}_z^{(j_1)} \otimes \hat{S}_z^{(j_2)} \otimes \dots \otimes \hat{S}_z^{(j_k)}$ . The number of such terms, and, thus, the signal strength, grows as  $N^k$ , but the quantum noise from the input states is unchanged. As a result, a sensitivity limit of  $\delta\mathcal{X} \propto N^{-k}$  applies when entanglement is used, and  $\delta\mathcal{X} \propto N^{-(k-1/2)}$  in the absence of entanglement<sup>7</sup>. For  $k \geq 2$ , this gives scaling better than the Heisenberg limit, so-called super-Heisenberg scaling<sup>7</sup>. We note that interactions and entanglement are compatible and both improve the scaling. The predicted advantage applies generally to quantum interferometry, and proposed mechanisms to produce metrologically relevant interactions include Kerr nonlinearities<sup>18</sup>, cold collisions in condensed atomic gases<sup>7</sup>, Duffing nonlinearity in nanomechanical resonators<sup>19</sup> and a two-pass effective nonlinearity with an atomic ensemble<sup>20</sup>. Topological excitations in nonlinear systems may also give advantageous scaling<sup>21</sup>.

In this Letter, we study interaction-based quantum metrology using unentangled probe particles. One challenge in demonstrating super-Heisenberg scaling is to engineer a suitable nonlinear Hamiltonian. Some nonlinearities have been shown to be intrinsically noisy<sup>14</sup> whereas others give super-Heisenberg scaling but fall short of the ideal,  $\delta\mathcal{X} \propto N^{-(k-1/2)}$ , under realistic conditions<sup>7,22</sup>. We use a cold atomic ensemble as a light–matter quantum interface<sup>12</sup> to produce quantum-noise-limited interactions, and use a Hamiltonian of the form  $\hat{H} = \hbar\mathcal{X}\hat{S}_z\hat{S}_0 = \hbar\mathcal{X}\hat{S}_zN/2$ . This Hamiltonian gives a polarization rotation that increases with photon number, without increasing quantum noise<sup>7</sup>. The experiment, shown schematically in Fig. 1, uses pulses of near-resonant light to measure the collective spin,  $\hat{\mathbf{F}}$ , of an ensemble of  $N_A \approx 10^6$  cold rubidium-87 atoms, probed on the  $5S_{1/2} \rightarrow 5P_{3/2}$  D<sub>2</sub> line. The experimental system is described in detail in refs 8, 23. The on-axis atomic magnetization,  $\langle \hat{F}_z \rangle$ , which plays the role of  $\mathcal{X}$  in this measurement, is prepared in the initial state  $\langle \hat{F}_z \rangle = N_A$  by optical pumping with resonant, circularly polarized light propagating along the trap axis,  $z$ . A weak, on-axis magnetic field is applied to preserve  $\hat{F}_z$  during the measurements.

Pulses of  $\hat{S}_x$ -polarized, but not entangled, photons pass through the ensemble and experience an optical rotation proportional to  $\langle \hat{F}_z \rangle$ . The light–atom interaction Hamiltonian  $\hat{H}_{\text{eff}} = \alpha^{(1)}\hat{F}_z\hat{S}_z + \beta^{(1)}\hat{F}_z\hat{S}_zN/2$  describes this paramagnetic Faraday rotation<sup>13</sup>. Both the linear term,  $\alpha^{(1)}\hat{F}_z\hat{S}_z$ , and the nonlinear term,  $\beta^{(1)}\hat{F}_z\hat{S}_zN/2$ , cause rotation of the plane of polarization from  $\hat{S}_x$  (vertical) towards  $\hat{S}_y$  (diagonal). Detection of  $\hat{S}_y$  then allows estimation of  $\hat{F}_z$ . As described in Supplementary Information,  $\alpha^{(1)}$  and  $\beta^{(1)}$  depend on the optical detuning,  $\Delta$ , relative to the  $F = 1 \rightarrow F' = 0$  transition; in particular,  $\alpha^{(1)}(\Delta_0) = 0$

<sup>1</sup>ICFO-Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain. <sup>2</sup>Laboratoire Matériaux et Phénomènes Quantiques, Université Paris Diderot et CNRS, UMR 7162, Bâtiment Condorcet, 75205 Paris Cedex 13, France.





**Figure 1 | Atom–light interface.** **a**, Experimental schematic: an ensemble of  $7 \times 10^5$   $^{87}\text{Rb}$  atoms, held in an optical dipole trap, is prepared in the state  $|F=1, m_F=1\rangle$  by optical pumping (OP). Linear ( $P_1, P_2$ ) and nonlinear ( $P_{\text{NL}}$ ) Faraday rotation probe pulses (in the order  $P_1, P_{\text{NL}}, P_2$ ) measure the atomic

for the specific detuning  $\Delta_0 \approx 2\pi \times 468.5$  MHz, allowing a purely non-linear estimation to be studied.

The rotation angle is  $\phi = \langle \hat{F}_z \rangle [A(\Delta) + B(\Delta)N]/2$ , where  $A \propto \alpha^{(1)}$  and  $B \propto \beta^{(1)}$  both account for the temporal pulse shape and the geometric overlap between the atomic density and the spatial mode of the probe. The shot-noise-limited uncertainty in the rotation angle, due to quantum uncertainty in the initial angle, is  $\delta\phi = N^{-1/2}$ . A contribution,  $\langle \hat{F}_z \rangle B(\Delta)\delta N/2$ , from initial number fluctuations  $\delta N = \langle N \rangle^{-1/2}$  is negligible for small rotation angles. This gives a measurement uncertainty of

$$\delta F_z = \langle \hat{F}_z \rangle \frac{\delta\phi}{\phi} = \frac{1}{A(\Delta)N^{1/2} + B(\Delta)N^{3/2}} \quad (1)$$

indicating a transition from SQL scaling,  $\delta F_z \propto N^{-1/2}$ , to super-Heisenberg scaling  $\delta F_z \propto N^{-3/2}$  with increasing  $N$ .

We use two probing regimes. The ‘linear probe’ consists of 40 1- $\mu\text{s}$  pulses (total illumination time,  $\tau_L = 40 \mu\text{s}$ ) spread over 400  $\mu\text{s}$  with detuning  $\Delta_L \gg \Delta_0$ . Together with the number of photons,  $N_L$ , used in the experiment for the linear probe, this gives  $A \gg N_L B$ , that is, linear estimation, and provides<sup>8</sup> a projection-noise-limited quantum non-demolition measurement<sup>24</sup> of  $\hat{F}_z$ , with uncertainty at the part-per-thousand level<sup>8</sup>. The ‘nonlinear probe’ consists of a single, Gaussian-shaped, high-intensity pulse with a full-duration at half-maximum of  $\tau_{\text{NL}} = 54$  ns,  $N_{\text{NL}}$  photons and a detuning  $\Delta_0$ , such that  $A \ll N_{\text{NL}} B$ . Crucially, having two probes allows us to calibrate the nonlinear measurement precisely using a highly sensitive and well-characterized independent measurement of the same sample.

We probe the same sample three times for each preparation. First we use the linear probe, which gives a precise and non-destructive measurement of  $\langle \hat{F}_z \rangle$  via the rotation angle,  $\phi_L$ . Then we use the nonlinear probe, which gives a rotation angle,  $\phi_{\text{NL}}$ , that is calibrated against the ‘true’ value (that is, with negligible error) provided by the linear probe. Finally we use a second linear probe to estimate the rotation angle  $\phi'_L$ , with which we can estimate the damage to the atomic magnetization,  $\eta \equiv 1 - \phi'_L/\phi_L$ , caused by the nonlinear probe.

The linear probe is calibrated using quantitative absorption imaging to measure  $N_A$ , and we find that  $A(\Delta_L) = 3.3(1) \times 10^{-8}$  rad per atom. The calibration of the nonlinear probe against the first linear probe is shown in Fig. 2: We repeat the above pump–probe sequence while varying  $N_A$  in the range  $1.5 \times 10^5$  to  $3.5 \times 10^5$  to generate a  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plot for a given value of  $N_{\text{NL}}$ . Because both  $\phi_L$  and  $\phi_{\text{NL}}$  are linear in  $N_A$ , we use linear regression to find the slope,  $b = d\phi_{\text{NL}}/d\phi_L = B(\Delta_0)N_{\text{NL}}/A(\Delta_L)$ , for that value of  $N_{\text{NL}}$ . The experiment is repeated for a range of different  $N_{\text{NL}}$  values.

The observed plot of  $b$  versus  $N_{\text{NL}}$ , shown in Fig. 2a, is well fitted by a simple model including saturation of the nonlinear response:

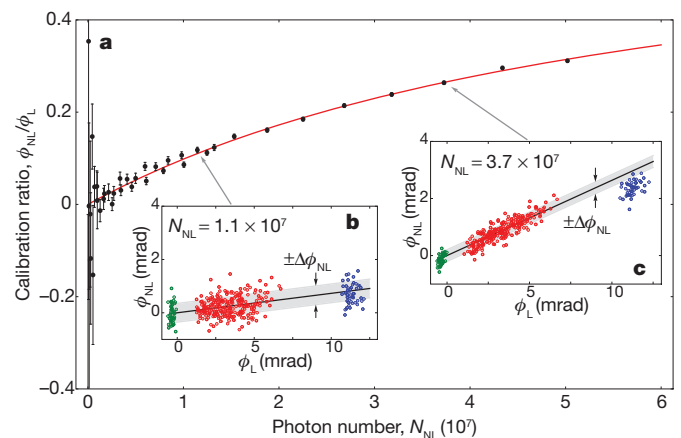
$$\frac{d\phi_{\text{NL}}}{d\phi_L} = \frac{B(\Delta_0)N_{\text{NL}}}{A(\Delta_L)} \frac{1}{1 + N_{\text{NL}}/N_{\text{NL}}^{(\text{sat})}} \quad (2)$$

magnetization, detected by a shot-noise-limited polarimeter (PM). The atom number is measured by quantitative absorption imaging (AI). **b**, Spectral positions of the pump, probe and imaging light on the  $D_2$  transition.

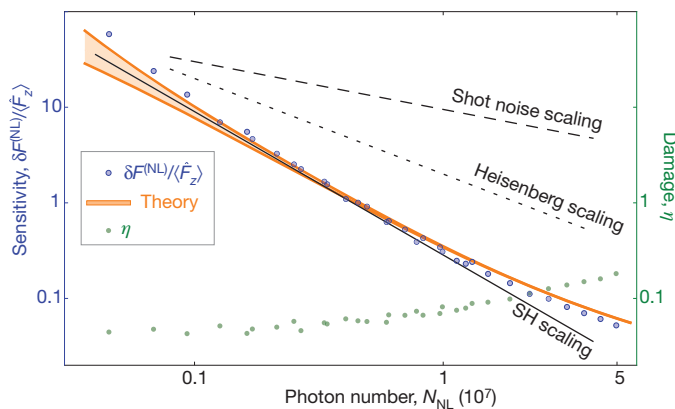
Here  $N_{\text{NL}}^{(\text{sat})} = 6.0(8) \times 10^7$  is a saturation parameter and the non-linear coupling strength is  $B(\Delta_0) = 3.8(2) \times 10^{-16}$  rad per atom per photon.

The noise in the nonlinear probe, again as a function of  $N_{\text{NL}}$ , is determined from the  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plots. As illustrated in Fig. 2b, c, the residual standard deviation of the fits indicates the observed uncertainty,  $\Delta\phi_{\text{NL}}$ , which includes the intrinsic uncertainty,  $\delta\phi_{\text{NL}}$ , and a small contribution from electronic noise. In Fig. 3, we plot the fractional sensitivity,  $\delta F_z^{(\text{NL})}/\langle \hat{F}_z \rangle$ , versus  $N_{\text{NL}}$ , calculated using equation (2) and considering the whole polarized ensemble, with  $\langle \hat{F}_z \rangle = 7 \times 10^5$  spins. In agreement with equation (1), the log–log slope indicates the scaling  $\delta F_z^{(\text{NL})} \propto N_{\text{NL}}^{-3/2}$  to within experimental uncertainties in the range  $N_{\text{NL}} = 10^6$  to  $10^7$ , and super-Heisenberg scaling, that is, steeper than  $N^{-1}$ , over two orders of magnitude ( $N_{\text{NL}} = 5 \times 10^5$  to  $5 \times 10^7$ ).

Results of numerical modelling using the Maxwell–Bloch equations to describe the nonlinear light–atom interaction are also shown in Fig. 3. Two curves are shown, for detunings  $\Delta_0 \pm (2\pi \times 200 \text{ kHz})$ , covering the combined uncertainty in  $\Delta$  due to the probe laser linewidth and inhomogeneous light shifts in the optical dipole trap. As expected from equation (1), this alters the sensitivity only at low  $N_{\text{NL}}$  values. The model is described in detail in Supplementary Information.



**Figure 2 | Calibration of nonlinear Faraday rotation.** **a**, Ratio of the nonlinear rotation,  $\phi_{\text{NL}}$ , to the linear rotation,  $\phi_L$ , versus the nonlinear probe photon number,  $N_{\text{NL}}$ . The data points and error bars indicate best-fit and standard errors from a linear regression,  $\phi_{\text{NL}} = b\phi_L + \text{const.}$ , for given values of  $N_{\text{NL}}$ . The red curve is a fit using equation (2), showing the expected nonlinear behaviour,  $\phi_{\text{NL}} \propto N_{\text{NL}}$ , with some saturation for large values of  $N_{\text{NL}}$ . **b, c**,  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plots for two values of  $N_{\text{NL}}$ . The atom number,  $N_A$ , is varied to produce a range of  $\phi_L$  and  $\phi_{\text{NL}}$  values. Green, no atoms ( $N_A = 0$ ); red,  $1.5 \times 10^5 < N_A < 3.5 \times 10^5$ ; blue,  $N_A \approx 7 \times 10^5$ . The blue circles are shown as a check on detector saturation, and are not included in the analysis.



**Figure 3 | Super-Heisenberg scaling.** Fractional sensitivity,  $\delta F_z^{(NL)}/\langle \hat{F}_z \rangle$ , of the nonlinear probe plotted versus the number of interacting photons,  $N_{NL}$ . Blue circles indicate the measured sensitivity, orange curves show results of numerical modelling, and the black lines indicate SQL, Heisenberg-limit and super-Heisenberg (SH) scaling for reference. Scaling surpassing the Heisenberg limit,  $\propto N_{NL}^{-1}$ , is observed over two orders of magnitude. The measured damage to the magnetization,  $\eta$ , shown as green circles, confirms the non-destructive nature of the measurement. Error bars for standard errors would be smaller than the symbols and are not shown.

For photon numbers above  $N_{NL} \approx 2 \times 10^7$ , the saturation of the nonlinear rotation alters the slope. This can be understood as optical pumping of atoms into states other than  $|F=1, m_F=1\rangle$  by the nonlinear probe. The damage to the atomic magnetization,  $\eta = 1 - \phi_L/\phi_L$ , also shown in Fig. 3, remains small, confirming the non-destructive nature of the measurement. The finite damage even for small  $N_{NL}$  values is possibly due to stray light and/or magnetic fields disturbing the atoms during the 20-ms period between the two linear measurements. For large  $N$ , higher-order nonlinear effects including optical pumping limit the range of super-Heisenberg scaling.

The experimental results illustrate the subtle relationship between scaling and sensitivity in a nonlinear system. For an ideal nonlinear measurement, the improved scaling would guarantee better absolute sensitivity for sufficiently large  $N$  values. Indeed, when the measurement bandwidth is taken into account, the nonlinear probe overtakes the linear one at  $N = 3.2 \times 10^6$ , where both achieve a sensitivity of  $1.1 \times 10^2$  spins  $\text{Hz}^{-1/2}$ . As a consequence, the nonlinear technique performs better in fast measurements. In contrast, when measurement time is not a limited resource, the comparison can be made on a 'sensitivity-per-measurement' basis and the ideal crossover point, of  $3.2 \times 10^3$  spins at  $N = 8.7 \times 10^7$ , is never actually reached, owing to the higher-order nonlinearities. Evidently, super-Heisenberg scaling allows but does not guarantee enhanced sensitivity: for the nonlinear technique to overtake the linear, it is also necessary that the scaling extend to large enough values of  $N$ . This example shows also that resource constraints dramatically influence the comparison between the linear and nonlinear techniques. See also Supplementary Information.

We have experimentally realized a system designed to achieve metrological sensitivity beyond the Heisenberg limit,  $\delta \mathcal{A} \propto N^{-1}$ , using metrologically relevant interactions among particles. To generate pairwise photon-photon interactions, we use fast, nonlinear optical effects in a cold atomic ensemble and measure the ensemble magnetization,  $\langle \hat{F}_z \rangle$ , with super-Heisenberg sensitivity  $\delta F_z \propto N^{-3/2}$ . To quantify the photon-photon interaction and the sensitivity rigorously, we calibrate against a precise, non-destructive, linear measurement of the same atomic quantity<sup>8</sup>, demonstrate quantum-noise-limited performance of the optical instrumentation and place an upper limit on systematic, that is, non-atomic, nonlinearities at the level of a few per cent. The experiment demonstrates the use of interparticle interactions as a new resource for quantum metrology. Although possible applications to precision measurement will require detailed study, our experiment

shows that interactions can produce super-Heisenberg scaling and improved precision in a quantum-limited measurement.

## METHODS SUMMARY

**Linear and nonlinear probe light.** The probe beam is aligned on the axis of the trap with a waist of 20  $\mu\text{m}$ , chosen to match the radial dimension of the cloud. In the linear probing regime, we use a train of 40 1- $\mu\text{s}$  pulses, with a repetition rate of 100 kHz, each containing  $3 \times 10^6$  photons detuned by +1.5 GHz from the  $F=1 \rightarrow F'=0$  transition. The maximum intensity is  $0.1 \text{ W cm}^{-2}$ . The signals are summed and can be considered a single, modulated pulse.

The nonlinear probe consists of a single, Gaussian-shaped pulse with a full-duration at half-maximum of 54 ns. The maximum intensity of the nonlinear probe is  $7 \text{ W cm}^{-2}$  for a pulse with  $10^7$  photons. Theory predicts that  $\alpha^{(1)} = 0$  at a detuning of  $\Delta = 2\pi \times 462 \text{ MHz}$  in free space. This is modified by trap-induced light shifts, and we use the empirical value  $\Delta_0 = 2\pi \times 468.5 \text{ MHz}$ , which gives zero rotation at low probe intensity.

**Instrumental noise.** The instrumental noise is quantified by measuring  $\text{var}(\hat{S}_y)$  versus input photon number  $N$  (that is,  $N_L$  or  $N_{NL}$ ), in the absence of atoms, to find contributions from electronic noise ( $V^{(el)} \propto N^0$ ), shot noise ( $N^1$ ) and technical noise ( $\propto N^2$ ), as described in Supplementary Information. We find that the contributions from electronic noise to the linear ( $V_L^{(el)}$ ) and nonlinear ( $V_{NL}^{(el)}$ ) probes are  $3 \times 10^5$  and  $4 \times 10^5$  per pulse, respectively, and that the technical noise is negligible. The instrumentation is thus shot-noise-limited over the full range of  $N$  used in the experiment. The intrinsic rotation uncertainty of the nonlinear probe,  $\delta\phi_{NL}$ , is calculated from the measured  $\Delta\phi_{NL}$  as  $(\delta\phi_{NL})^2 = (\Delta\phi_{NL})^2 - V_{NL}^{(el)}$ . The correction is at most 5%.

**Instrumental linearity.** The linearity of the experimental system and analysis is verified by using a wave plate in place of the atoms to produce a linear rotation equal to the largest observed nonlinear rotation. Over the full range of photon numbers used in the experiment, the detected rotation angle is constant to within 5%, and SQL scaling is observed.

Received 31 July; accepted 15 December 2010.

- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).
- Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. Super-resolving phase measurements with a multiphoton entangled state. *Nature* **429**, 161–164 (2004).
- Lee, H., Kok, P. & Dowling, J. P. A quantum Rosetta stone for interferometry. *J. Mod. Opt.* **49**, 2325–2338 (2002).
- Boixo, S., Flammaria, S. T., Caves, C. M. & Geremia, J. Generalized limits for single-parameter quantum estimation. *Phys. Rev. Lett.* **98**, 090401 (2007).
- Choi, S. & Sundaram, B. Bose-Einstein condensate as a nonlinear Ramsey interferometer operating beyond the Heisenberg limit. *Phys. Rev. A* **77**, 053613 (2008).
- Roy, S. M. & Braunstein, S. L. Exponentially enhanced quantum metrology. *Phys. Rev. Lett.* **100**, 220501 (2008).
- Boixo, S. *et al.* Quantum metrology: dynamics versus entanglement. *Phys. Rev. Lett.* **101**, 040403 (2008).
- Koschorreck, M., Napolitano, M., Dubost, B. & Mitchell, M. W. Sub-projection-noise sensitivity in broadband atomic magnetometry. *Phys. Rev. Lett.* **104**, 093602 (2010).
- Koschorreck, M., Napolitano, M., Dubost, B. & Mitchell, M. W. Quantum nondemolition measurement of large-spin ensembles by dynamical decoupling. *Phys. Rev. Lett.* **105**, 093602 (2010).
- Kominis, I., Kornack, T., Allred, J. & Romalis, M. A subfemtotesla multichannel atomic magnetometer. *Nature* **422**, 596–599 (2003).
- Budker, D. & Romalis, M. Optical magnetometry. *Nature Phys.* **3**, 227–234 (2007).
- Hammerer, K., Sørensen, A. S. & Polzik, E. S. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.* **82**, 1041–1093 (2010).
- Napolitano, M. & Mitchell, M. W. Nonlinear metrology with a quantum interface. *N. J. Phys.* **12**, 093016 (2010).
- Fleischhauer, M., Matsuoka, A. B. & Scully, M. O. Quantum limit of optical magnetometry in the presence of ac Stark shifts. *Phys. Rev. A* **62**, 013808 (2000).
- Scully, M. O., Englert, B. G. & Walther, H. Quantum optical tests of complementarity. *Nature* **351**, 111–116 (1991).
- Wasilewski, W. *et al.* Quantum noise limited and entanglement-assisted magnetometry. *Phys. Rev. Lett.* **104**, 133601 (2010).
- Wolffgramm, F. *et al.* Squeezed-light optical magnetometry. *Phys. Rev. Lett.* **105**, 053601 (2010).
- Beltrán, J. & Luis, A. Breaking the Heisenberg limit with inefficient detectors. *Phys. Rev. A* **72**, 045801 (2005).
- Woolley, M. J., Milburn, G. J. & Caves, C. M. Nonlinear quantum metrology using coupled nanomechanical resonators. *N. J. Phys.* **10**, 125018 (2008).
- Chase, B. A., Baragiola, B. Q., Partner, H. L., Black, B. D. & Geremia, J. M. Magnetometry via a double-pass continuous quantum measurement of atomic spin. *Phys. Rev. A* **79**, 062107 (2009).
- Negretti, A., Henkel, C. & Mølmer, K. Quantum-limited position measurements of a dark matter-wave soliton. *Phys. Rev. A* **77**, 043606 (2008).

22. Boixo, S. *et al.* Quantum-limited metrology and Bose-Einstein condensates. *Phys. Rev. A* **80**, 032103 (2009).
23. Kubasik, M. *et al.* Polarization-based light-atom quantum interface with an all-optical trap. *Phys. Rev. A* **79**, 043815 (2009).
24. Braginskii, V. B. & Vorontsov, Y. I. Quantum-mechanical limitations in macroscopic experiments and modern experimental technique. *Sov. Phys. Usp.* **17**, 644 (1975).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank I. H. Deutsch and F. Illuminati for comments. We thank C. M. Caves and A. D. Codorníu for inspiration. This work was supported by the Spanish

Ministry of Science and Innovation through the Consolider-Ingenio 2010 project QOIT, the Ingenio-Explora project OCHO (ref. FIS2009-07676-E/FIS) and project ILUMA (ref. FIS2008-01051), by the Marie-Curie RTN EMALI, and by Fundacio CELLEX Barcelona.

**Author Contributions** All authors contributed equally to the work presented in this paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.N. ([mario.napolitano@icfo.es](mailto:mario.napolitano@icfo.es)).



# Collapse of long-range charge order tracked by time-resolved photoemission at high momenta

Timm Rohwer<sup>1\*</sup>, Stefan Hellmann<sup>1\*</sup>, Martin Wiesenmayer<sup>1</sup>, Christian Sohrt<sup>1</sup>, Ankatrin Stange<sup>1</sup>, Bartosz Slomski<sup>1</sup>, Adra Carr<sup>2</sup>, Yanwei Liu<sup>3,4</sup>, Luis Miaja Avila<sup>5</sup>, Matthias Kalläne<sup>1</sup>, Stefan Mathias<sup>2,6</sup>, Lutz Kipp<sup>1</sup>, Kai Rossnagel<sup>1</sup> & Michael Bauer<sup>1</sup>

Intense femtosecond ( $10^{-15}$  s) light pulses can be used to transform electronic, magnetic and structural order in condensed-matter systems on timescales of electronic and atomic motion<sup>1,2,3</sup>. This technique is particularly useful in the study<sup>4,5</sup> and in the control<sup>6</sup> of materials whose physical properties are governed by the interactions between multiple degrees of freedom. Time- and angle-resolved photoemission spectroscopy is in this context a direct and comprehensive, energy- and momentum-selective probe of the ultrafast processes that couple to the electronic degrees of freedom<sup>7–10</sup>. Previously, the capability of such studies to access electron momentum space away from zero momentum was, however, restricted owing to limitations of the available probing photon energy<sup>10,11</sup>. Here, using femtosecond extreme-ultraviolet pulses delivered by a high-harmonic-generation source, we use time- and angle-resolved photoemission spectroscopy to measure the photoinduced vaporization of a charge-ordered state in the potential excitonic insulator 1T-TiSe<sub>2</sub> (refs 12, 13). By way of stroboscopic imaging of electronic band dispersions at large momentum, in the vicinity of the edge of the first Brillouin zone, we reveal that the collapse of atomic-scale periodic long-range order happens on a timescale as short as 20 femtoseconds. The surprisingly fast response of the system is assigned to screening by the transient generation of free charge carriers. Similar screening scenarios are likely to be relevant in other photoinduced solid-state transitions and may generally determine the response times. Moreover, as electron states with large momenta govern fundamental electronic properties in condensed matter systems<sup>14</sup>, we anticipate that the experimental advance represented by the present study will be useful to study the ultrafast dynamics and microscopic mechanisms of electronic phenomena in a wide range of materials.

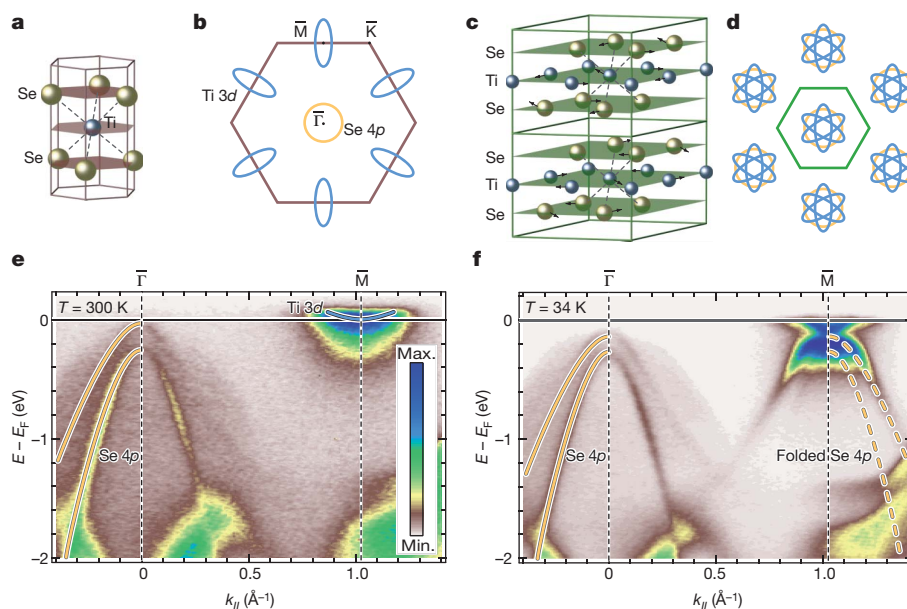
The electronic properties of many condensed matter systems are determined by large-momentum electron states, often located near the edge of the first Brillouin zone (BZ), the unit cell of crystalline solids in electron momentum ( $k$ ) space. A prominent current example is graphene, whose hallmark, the critical crossing point of its peculiar conical band dispersion, is located at the corner of the BZ<sup>15</sup>. The copper oxide based high-temperature superconductors represent another example, where the much debated competition between the pseudogap and superconductivity is particularly pronounced in the so-called antinodal region of the Fermi surface, near the BZ boundary<sup>16</sup>. A final example is the new class of iron-based superconductors, which are characterized by magnetic excitations that couple two sets of Fermi surfaces, one set centred on the corners of the BZ<sup>17</sup>. More generally, it is the coupling of high momentum electron states near the Fermi momenta  $k_F$  (the momenta of the highest occupied electron states) that contributes most to the linear response function of an electron liquid, sometimes even causing a divergence that leads to phase instabilities.

Typical electron momenta near the boundary of the first BZ are in the  $1 \text{ \AA}^{-1}$  regime. Conventional angle-resolved photoemission spectroscopy (ARPES) with photon energies exceeding roughly 10 eV is probably the most powerful tool to map band structure peculiarities and Fermi surfaces up to and beyond these critical points: this technique can also, at the same time, determine many-body effects embodied in the fine details of band dispersions and in the distribution of spectral weight. ARPES is particularly well suited for layered materials—as in the present study—because for quasi-two-dimensional systems, the measured electronic structure can be considered as being predominantly characteristic of the bulk, despite the surface sensitivity of the probe. The great allure of corresponding time-resolved ARPES experiments is the provision of direct dynamical information and the possibility of disentangling—via temporal discrimination—the various interactions between the relevant degrees of freedom that determine material properties in the quantum world. Here, we present femtosecond time-resolved ARPES experiments, in which transient changes in the whole occupied electronic structure between the centre and the edge of the BZ are probed to answer a simple fundamental question: how fast can long-range charge order in a solid melt?

The charge-ordered state we investigate is the conspicuous ( $2 \times 2 \times 2$ ) charge-density wave (CDW) that occurs in the layered compound 1T-TiSe<sub>2</sub> below a temperature of 200 K (ref. 13). Figure 1 presents the thermal equilibrium view of the CDW transition, which affects both structural and electronic properties. On the transition, the atoms move to new equilibrium positions such that the real-space unit cell doubles its size in all three directions (Fig. 1a and c; arrows in Fig. 1c indicate the atomic displacements from the normal-phase positions); in momentum space, the dimensions of the BZ are correspondingly halved (Fig. 1b and d). The new  $k$ -space geometry suggests that the wave vector of the CDW is determined by an interaction between the Se 4p valence band maximum at  $\bar{\Gamma}$ , the centre of the BZ, and the elliptical pocket of Ti 3d states at the BZ edge at  $\bar{M}$  (Fig. 1b). These symmetry points are connected by the new reciprocal lattice vectors and become equivalent  $\bar{\Gamma}$  points in the new phase, which allows direct Se 4p–Ti 3d interaction (Fig. 1d). This interaction is in fact remarkably strong and extremely well resolved by conventional ARPES<sup>12,18,19</sup>. Figure 1e and f compares ARPES intensity maps recorded with synchrotron radiation ( $h\nu = 119 \text{ eV}$ ) along the  $\bar{\Gamma}-\bar{M}$  direction above and below the CDW transition temperature. At  $\bar{\Gamma}$ , the downward dispersing (hole-like) Se 4p bands dominate the photoemission signal in both maps and only a small shift of the valence band maximum is visible. At  $\bar{M}$ , where the high-temperature map shows the bottom of the upward dispersing (electron-like) Ti 3d band, the effects are more dramatic: the CDW leads to a strong selective transfer of spectral weight by folding the Se 4p band from  $\bar{\Gamma}$  onto  $\bar{M}$ . It is this remarkably strong folded Se 4p intensity which we will use in the time-resolved experiments as a spectroscopic measure for CDW order in 1T-TiSe<sub>2</sub>.

<sup>1</sup>Institute of Experimental and Applied Physics, University of Kiel, 24118 Kiel, Germany. <sup>2</sup>JILA, University of Colorado and NIST, Boulder, Colorado 80309-0440, USA. <sup>3</sup>University of California, Berkeley, NSF ERC Extreme Ultraviolet Science and Technology, Berkeley, California 94720, USA. <sup>4</sup>Center for X-Ray Optics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. <sup>5</sup>Center for Nano and Molecular Science, University of Texas at Austin, Austin, Texas 78722, USA. <sup>6</sup>Department of Physics and Research Center OPTIMAS, University of Kaiserslautern, 67663 Kaiserslautern, Germany.

\*These authors contributed equally to this work.



**Figure 1 | CDW phase transition of 1T-TiSe<sub>2</sub>.** **a**, Real-space unit cell of the normal phase at room temperature. **b**, Common unit cell (first BZ, purple line) in momentum space of the room temperature phase (two-dimensional projection onto the surface plane).  $\bar{\Gamma}$ ,  $\bar{M}$  and  $\bar{K}$  are high-symmetry points of the first BZ. Blue ellipses and yellow circle indicate the Fermi surface topology of Ti 3d and Se 4p bands, respectively. Planes show the atomic layers of the two-dimensional crystalline structure. **c**, Real-space unit cell of the CDW phase.

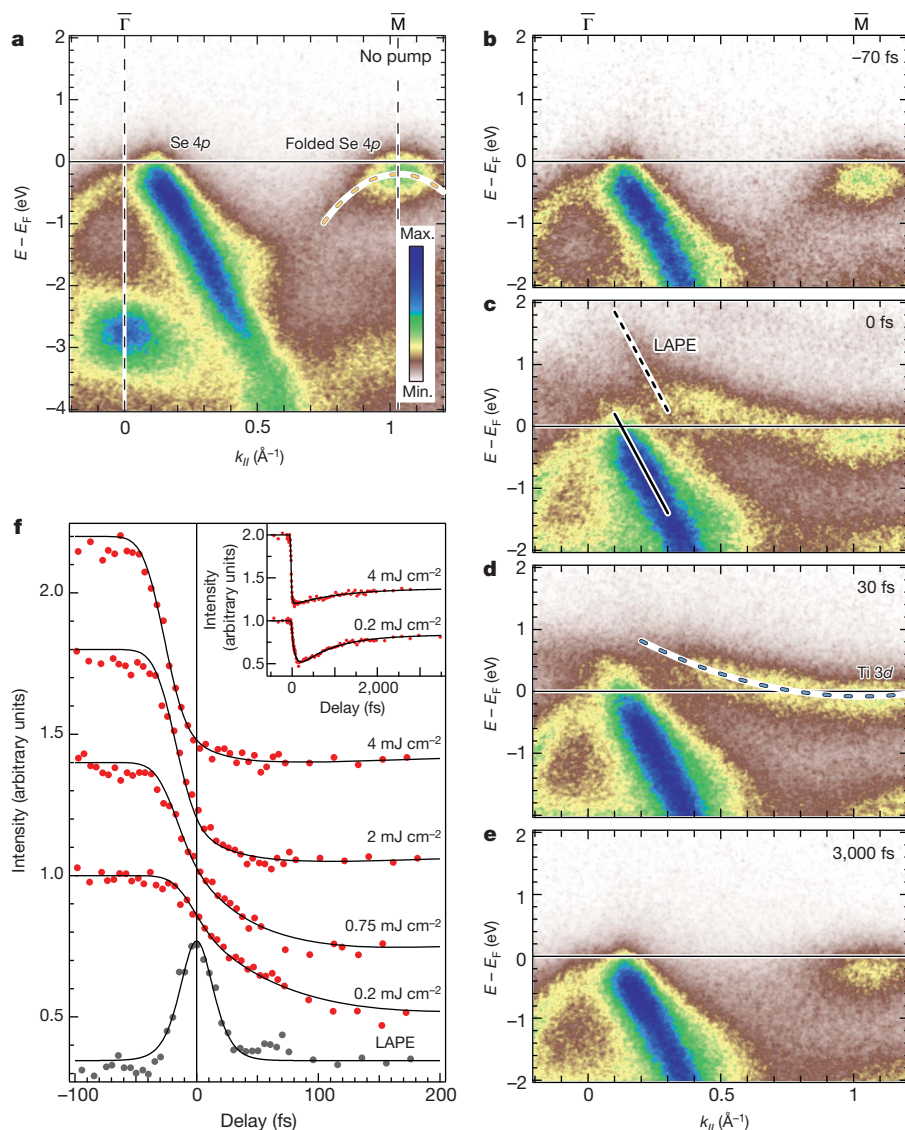
We now apply sub-10-fs extreme ultraviolet (XUV) pulses ( $h\nu = 43$  eV, s-polarization) to monitor the transient response of the CDW phase to excitation with infrared laser pulses ( $h\nu = 1.57$  eV) of 32 fs width. Details of the time-resolved ARPES experiment are described in Methods. Figure 2a shows ARPES intensity maps of 1T-TiSe<sub>2</sub> measured at  $T = 125$  K with the femtosecond XUV light source<sup>20</sup> (Methods). Despite the poorer energy resolution, both sets of Se 4p bands are well resolved, the original one at  $\bar{\Gamma}$  and the folded one at  $\bar{M}$ . Time-resolved experiments are performed at infrared pump fluences between 0.2 and 5 mJ cm<sup>-2</sup>, corresponding to an excitation density range of 0.025 to 0.63 photons per Ti atom. Figure 2b–e shows four photoemission snapshots recorded at a pump fluence of 5 mJ cm<sup>-2</sup> with increasing temporal delay between the infrared pump and the XUV probe, up to a maximum of 3 ps (Supplementary Movie 1). The data have been corrected for a space charge shift of 200 meV induced by the electron background because of multi-photon photoemission by the infrared pump pulse. The time series is dominated by two prominent changes in the photoemission intensity maps. First, in instantaneous response to the infrared excitation, an electron-like band appears, crossing the Fermi energy  $E_F$  and extending (at sufficiently small temporal delays) from  $\bar{M}$  to  $\bar{\Gamma}$ . We observe here the transient generation of quasi-free charge carriers because of near-resonant Ti 3d – Se 4p excitation. Second, the downward-dispersing Se 4p band, folded onto the  $\bar{M}$  point owing to the interaction with the CDW superlattice, disappears or is at least considerably reduced in intensity. This suggests that long-range order in the electronic subsystem breaks down on an ultrafast timescale. In the following, we restrict our quantitative analysis to the short-time (sub-100-fs) dynamics of this process.

Figure 2f compares the temporal evolution of the integrated intensity of the folded Se 4p band—our spectroscopic measure for CDW order—for different pump fluences (see Supplementary Information section 3 for details of the data analysis). Both breakdown and (partial) recovery of the signal (inset of Fig. 2f) are strongly dependent on the pump fluence. The fluence dependence of the time constant characterising the signal breakdown,  $\tau_{\text{Se } 4p}$ , is shown in Fig. 3a: at the lowest fluences,

Arrows indicate the atomic displacements from the normal-phase positions; the dashed lines indicate the extension of the unit cell in the normal phase. **d**, First BZ (green line) of the CDW phase. The folding of Se 4p and Ti 3d states is indicated. **e**, ARPES intensity map (electron binding energy versus momentum) of the room temperature phase. Photoelectron intensity is encoded in a false-colour scale. **f**, ARPES intensity map of the CDW phase.

the initial drop in the signal is retarded by about 80 fs with respect to the laser pulse excitation. As the fluence increases, the response becomes continuously faster, and at the highest fluences the transient minimum in the folded Se 4p band intensity appears well within the 32-fs-long infrared pump laser pulse with an ultimate response time of 20 fs. For comparison, the dynamics associated with the initial population of the Ti 3d band due to absorption happens within the width of the infrared pulse for the entire pump fluence regime. Notably, for the highest excitation fluence, the folded Se 4p band follows this population dynamics without delay (Supplementary Information section 4). The partial recovery of the folded Se 4p intensity is observed on timescales of several hundreds of femtoseconds. Two-temperature model calculations following reference 21 suggest that this recovery is mostly driven by thermalization of the electronic subsystem with the atomic lattice.

In previous studies, it has been shown that the fundamental time-scales of photoinduced phase transitions are governed by bottlenecks associated with the characteristic response times of the relevant degrees of freedom, such as the oscillation period of neighbouring atoms<sup>22</sup> or the hopping rate of localized electrons between neighbouring sites<sup>23</sup>. The upper solid line in Fig. 3a marks for instance the expected short-time limit (75 fs) of the lattice response of 1T-TiSe<sub>2</sub> to a photoexcitation; this short-time limit is taken as one-quarter of the oscillation period of the high-frequency CDW amplitude mode<sup>24</sup>. The vaporization of a long-range-ordered state within 20 fs is in this context exceptionally fast. The ultrafast timescale observed in the high-fluence regime points to a purely electronically driven process, whose response time, however, strongly depends on the excitation fluence. The absorption of the light pulse initially increases the free charge carrier density  $n$  (electrons and holes), as can be seen in the instantaneous population of the Ti 3d band. This transient free carrier population, which is directly governed by the excitation fluence, links the time constant  $\tau_{\text{Se } 4p}$  to a material specific timescale: quantum kinetic calculations have shown that the characteristic build-up time for carrier screening in response to an ultrashort laser excitation is the plasma oscillation period,  $\tau_{\text{pl}}$  (refs 25, 26), which scales with  $1/\sqrt{n}$ . As shown in Fig. 3a,  $\tau_{\text{Se } 4p}$  closely follows such a  $1/\sqrt{n}$  dependence (a quantitative estimate of the



**Figure 2 | Tracking the photoinduced transition by femtosecond time-resolved ARPES.** **a**, ARPES intensity map of 1T-TiSe<sub>2</sub> recorded with high-harmonic XUV pulses in the CDW phase (temperature  $T = 125$  K). **b–e**, Time-resolved ARPES snapshots at increasing pump (infrared)–probe (XUV) temporal delays (pump–pulse fluence,  $5 \text{ mJ cm}^{-2}$ ). Energy distribution curves (EDCs) are provided in Supplementary Information 2. **f**, Photoemission transients of the folded Se 4*p* band for different infrared excitation fluences close to time zero. The infrared–XUV cross-correlation signal, which has been determined from the laser assisted photoemission<sup>30</sup> (LAPE) signal, is added (Methods). Inset, transients up to temporal delays of 3.5 ps.

photoinduced carrier density  $n$  and the corresponding plasma oscillation period is given in Supplementary Information section 5). It has in fact been shown that the build-up of screening by photo-injected carriers is relevant for the dynamics of ultrafast processes on the sub-100-fs timescale<sup>27</sup>.

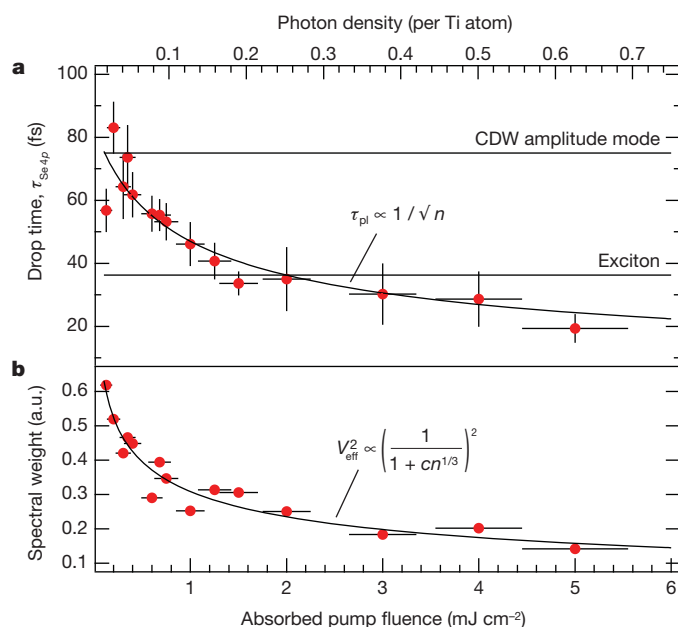
Suppression of screening is important in correlation-induced phase transitions. Therefore, it is not surprising that the reverse—the enhancement of screening by photo-injection of free carriers—can destroy a correlation-induced phase. The CDW transition in 1T-TiSe<sub>2</sub> has repeatedly been associated with an excitonic insulator instability<sup>12,13</sup>. In this model, the transition is driven by the spontaneous formation of excitons, which can occur in semiconductors or semimetals when the bandgap or band overlap becomes smaller than the exciton binding energy. The exciton formation requires a low concentration of mobile charge carriers and a correspondingly poorly screened Coulomb interaction. In the case of 1T-TiSe<sub>2</sub>, the narrow gap/overlap system would become unstable to the formation of Ti 3*d* – Se 4*p* excitons and would exhibit a new periodicity governed by the wave vector connecting the corresponding valence and conduction band pockets. This purely electronic process seems consistent with the measured ultrafast response times and a screening-based interpretation, as bound excitons would certainly be screened by the photo-injected carriers. For a transition temperature of 200 K, energy–time uncertainty yields a response time of 35 fs for such a purely excitonic process (see lower solid line in Fig. 3a). More generally, however, our results are in line with the

screening of any (unspecified) interaction between Se 4*p* and Ti 3*d* states described by an effective matrix element  $V_{\text{eff}}^2$ . The spectral weight of the folded Se 4*p* state will scale with  $V_{\text{eff}}^2$  and we can estimate the effect of screening within the Thomas–Fermi approach<sup>14</sup>. This yields  $V_{\text{eff}}^2 \approx (1 + \text{const.} \times n^{1/3})$ , in reasonable agreement with our experimental data (see Fig. 3b). A deeper analysis, particularly with respect to the puzzling story of the CDW phase transition in 1T-TiSe<sub>2</sub>, requires a sophisticated theoretical description which, for instance, considers the quantum kinetics of screening in non-equilibrium systems or effects arising from the photo-doping of excitonic insulators.

We finally address the question of to what extent the screening by the nascent carriers affects the properties of 1T-TiSe<sub>2</sub> on the 20-fs timescale. As discussed above, the response of the atomic lattice is slow so that structurally the sample will still resemble the CDW phase on this timescale. However, the valence electronic structure becomes substantially modified as soon as screening becomes effective. In fact, electronically, the system undergoes a transition from a poorly conducting CDW state into a metallic phase within 20 fs. This is not only implied by the screening scenario but is also directly visible in Fig. 2d, which shows the transient metallization of the Ti 3*d* band only 30 fs after the optical excitation.

The ultrafast breakdown of long-range charge order that we report here is much faster than the material-characteristic oscillations of collective modes that are commonly thought to limit the response times in photoinduced processes. This surprising result may therefore stimulate





**Figure 3 | Fluence dependence of the photoinduced transition.** **a**, Signal drop time,  $\tau_{\text{Se } 4p}$ , of the folded Se 4p band (filled red circles) as a function of the absorbed pump fluence. The  $\sim 10\%$  error bars in the absorbed pump fluence are determined from the uncertainty in the optical constants of 1T-TiSe<sub>2</sub>, the stability of the pump laser, and the uncertainty in the pump pulse diameter at the sample position. The error bars in the drop time indicate the standard deviation of the fits to the data. **b**, Minimum transient spectral weight in the folded Se 4p band (red dots) as a function of absorbed fluence. Equation of curve fitted to data is shown in each panel. a.u., arbitrary units.

new concepts for ultrafast switching devices. Furthermore, in our study we have essentially monitored the intensity of a superlattice Bragg peak in electron momentum space. Thus, from a methodological point of view, the presented time-resolved ARPES approach complements present-day time-resolved diffraction experiments<sup>28,29</sup>, with the advantage of an exceptionally high temporal resolution.

## METHODS SUMMARY

**Sample preparation.** 1T-TiSe<sub>2</sub> single crystals were grown from the elements by chemical vapour transport using iodine as transport agent. Before the photoemission measurements, the samples were cleaved *in situ* at room temperature in ultrahigh vacuum.

**Photoemission measurements.** Static ARPES experiments were conducted at beamline 7.0.1 of the Advanced Light Source at Berkeley with a Scienta R4000 electron spectrometer. The photon energy was 119 eV and the overall energy resolution was  $\sim 30$  meV. Femtosecond time-resolved ARPES measurements were conducted at the University of Kiel with a SPECS Phoibos 150 electron spectrometer. Here, the photon energy was 43 eV and the overall energy resolution in the experiment was  $\sim 400$  meV. The light source for the time-resolved experiments was an argon-filled hollow-fibre waveguide (XUUS, KMLabs) for high harmonic generation operated with a 3 kHz Ti:sapphire amplifier system (Dragon, KMLabs, pumped by an Empower 30, Spectra Physics).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 6 September 2010; accepted 11 January 2011.

Published online 9 March 2011.

- Yonemitsu, K. & Nasu, K. Theory of photoinduced phase transitions in itinerant electron systems. *Phys. Rep.* **465**, 1–60 (2008).
- Bargheer, M., Zhavoronkov, N., Woerner, M. & Elsaesser, T. Recent progress in ultrafast X-ray diffraction. *ChemPhysChem* **7**, 783–792 (2006).
- Bigot, J.-Y., Vomer, M. & Beaupre, E. Coherent ultrafast magnetism induced by femtosecond laser pulses. *Nature Phys.* **5**, 515–520 (2009).

- Chollet, M. *et al.* Gigantic photoresponse in  $1/4$ -filled-band organic salt (EDO-TTF)<sub>2</sub>PF<sub>6</sub>. *Science* **307**, 86–89 (2005).
- Demsar, J., Biljaković, K. & Mihailović, D. Single particle and collective excitations in the one-dimensional charge density wave solid K<sub>0.3</sub>MoO<sub>3</sub> probed in real time by femtosecond spectroscopy. *Phys. Rev. Lett.* **83**, 800–803 (1999).
- Rini, M. *et al.* Control of the electronic phase of a manganite by mode-selective vibrational excitation. *Nature* **449**, 72–74 (2007).
- Haight, R. & Silbermann, J. A. Surface intervalley scattering on GaAs(110): direct observation with picosecond laser photoemission. *Phys. Rev. Lett.* **62**, 815–818 (1989).
- Fann, W. S., Storz, R., Tom, H. W. K. & Bokor, J. Electron thermalization in gold. *Phys. Rev. B* **46**, 13592–13595 (1992).
- Bauer, M. *et al.* Direct observation of surface chemistry using ultrafast soft-X-ray pulses. *Phys. Rev. Lett.* **87**, 025501 (2001).
- Perfetti, L. *et al.* Time evolution of the electronic structure of 1T-TaS<sub>2</sub> through the insulator-metal transition. *Phys. Rev. Lett.* **97**, 067402 (2006).
- Schmitt, F. *et al.* Transient electronic structure and melting of a charge density wave in TbTe<sub>3</sub>. *Science* **321**, 1649–1652 (2008).
- Cercellier, H. *et al.* Evidence for an excitonic insulator phase in 1T-TiSe<sub>2</sub>. *Phys. Rev. Lett.* **99**, 146403 (2007).
- Di Salvo, F. J., Moncton, D. E. & Waszczak, J. V. Electronic properties and superlattice formation in the semimetal TiSe<sub>2</sub>. *Phys. Rev. B* **14**, 4321–4328 (1976).
- Ashcroft, N. W. & Mermin, N. D. *Solid State Physics* (Brooks/Cole, Belmont, 1976).
- Ohta, T., Bostwick, A., Seyller, T., Horn, K. & Rotenberg, E. Controlling the electronic structure of bilayer graphene. *Science* **313**, 951–954 (2006).
- Kondo, T., Khasanov, R., Takeuchi, T., Schmalian, J. & Kaminski, A. Competition between the pseudogap and superconductivity in the high-*T<sub>c</sub>* copper oxides. *Nature* **457**, 296–300 (2009).
- Mazin, I. Superconductivity gets an iron boost. *Nature* **464**, 183–186 (2010).
- Rosnagel, K., Kipp, L. & Skibowski, M. Charge-density-wave phase transition in 1T-TiSe<sub>2</sub>: excitonic insulator versus band-type Jahn-Teller mechanism. *Phys. Rev. B* **65**, 235101 (2002).
- Kidd, T. E., Miller, T., Chou, M. Y. & Chiang, T.-C. Electron-hole coupling and the charge density wave transition in TiSe<sub>2</sub>. *Phys. Rev. Lett.* **88**, 226402 (2002).
- Rundquist, A. *et al.* Phase matching of soft-X-ray harmonic emission in hollow-core fibers. *Science* **280**, 1412–1415 (1998).
- Anisimov, S. I., Kapeliovich, B. L. & Perel'man, T. L. Electron-emission from surface of metals induced by ultrashort laser pulses. *Sov. Phys. JETP* **39**, 375–377 (1974).
- Cavalleri, A. *et al.* Evidence for a structurally-driven insulator-to-metal transition in VO<sub>2</sub>: a view from the ultrafast timescale. *Phys. Rev. B* **70**, 161102(R) (2004).
- Wall, S. *et al.* Quantum interference between charge excitation paths in a solid-state Mott insulator. *Nature Phys.* advance online publication, doi:10.1038/nphys1831 (5 December 2010).
- Holy, J. A., Woo, K. C., Klein, M. V. & Brown, F. C. Raman and infrared studies of superlattice formation in TiSe<sub>2</sub>. *Phys. Rev. B* **16**, 3628–3637 (1977).
- El Sayed, K., Schuster, S., Haug, H., Herzel, F. & Henneberger, K. Subpicosecond plasmon response: buildup of screening. *Phys. Rev. B* **49**, 7337–7344 (1994).
- Bányai, L., Vu, Q. T., Mieck, B. & Haug, H. Ultrafast quantum kinetics of time-dependent RPA-screened Coulomb scattering. *Phys. Rev. Lett.* **81**, 882–885 (1998).
- Huber, R. *et al.* How many-particle interactions develop after ultrafast excitation of an electron-hole plasma. *Nature* **414**, 286–289 (2001).
- Sokolowski-Tinten, K. *et al.* Femtosecond X-ray measurement of coherent lattice vibrations near the Lindemann stability limit. *Nature* **422**, 287–289 (2003).
- Siwick, B. J., Dwyer, J. R., Jordan, R. E. & Miller, R. J. D. An atomic-level view of melting using femtosecond electron diffraction. *Science* **302**, 1382–1385 (2003).
- Majava-Avila, L. *et al.* Laser-assisted photoelectric effect from surfaces. *Phys. Rev. Lett.* **97**, 113604 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** M.B. and S.M. thank M. Aeschlimann for support and discussion. M.B. and S.M. also thank M. Murnane and H. Kapteyn for their support through the NSF EUV ERC. A.C. acknowledges support from the JILA Physics Frontier Center. This work was supported by the German Science Foundation (DFG) within the SFB 855 (C.S., M.B., L.K., K.R.) and by the European Community's FP7 under Marie Curie International Outgoing Fellowship GA 253316 (S.M.). Operation of the Advanced Light Source is supported by the US Department of Energy, Office of Basic Energy Sciences.

**Author Contributions** M.B. and K.R. conceived the experiment and wrote the paper. T.R., S.H., M.W., B.S., S.M., L.K., M.B. and K.R. realized the experimental time-resolved ARPES setup. A.C., L.M.A. and Y.L. designed and fabricated the EUV multilayer mirrors. T.R., S.H., M.W., C.S. and A.S. collected the time-resolved photoemission data and performed the data analysis. M.K. and K.R. collected and analysed the static photoemission data at the Advanced Light Source. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.B. (bauer@physik.uni-kiel.de; experimental technique) or K.R. (rosnagel@physik.uni-kiel.de; experimental data).

## METHODS

**Sample preparation and photoemission set-up.** 1T-TiSe<sub>2</sub> single crystals were grown from the elements by chemical vapour transport using iodine as transport agent. Samples were mounted on a cryogenic manipulator and cleaved *in situ* under ultrahigh vacuum (UHV) conditions at a base pressure of  $3 \times 10^{-10}$  mbar. Conventional ARPES measurements were conducted at beamline 7.0.1 of the Advanced Light Source at Berkeley, with a Scienta R4000 electron spectrometer. The photon energy was 119 eV and the overall energy resolution was  $\sim 30$  meV. Femtosecond time-resolved ARPES measurements were conducted at the University of Kiel. Photoemitted electrons were detected using a hemispherical electron energy analyzer (SPECS, Phoibos 150) equipped with a two-dimensional detection unit for parallel energy and momentum detection. The total energy resolution of the experiment was mainly governed by the spectral broadening of the femtosecond XUV pulses and was determined to be  $\sim 400$  meV. An independent characterization of the high harmonic radiation with a grating spectrometer showed that the spectral width of the used 27th harmonic was  $340 \pm 40$  meV. Typical integration times for analysis-grade spectra were 3 min. High-quality data as shown in Fig. 2 required an integration time of 15 min.

**Pulsed XUV light source.** The light source used for the pump–probe photoemission experiments was a 3-kHz Ti:sapphire amplifier system (Dragon, KMLabs, pumped by an Empower 30, Spectra Physics) delivering infrared pulses at 790 nm, 1.2 mJ pulse energy and 32 fs pulse duration. For photoemission, 80% of the pulse energy was used to generate high harmonic femtosecond XUV pulses in an argon-filled hollow-fibre waveguide (XUUS, KMLabs). A pair of multilayer mirrors (total reflectivity, 13%) selected the 27th harmonic ( $h\nu = 43$  eV) out of the harmonic spectrum and focused it at an angle of  $45^\circ$  onto the sample mounted in the UHV system. The intensity of the 27th harmonic at the sample position was measured *in situ* using a calibrated XUV photodiode (SXUV 20 HS1, International Radiation Detections) yielding a fluence in the  $10^9$  photons  $\text{s}^{-1}$  regime. The residual 20% of the amplifier output was available for infrared photoexcitation of the 1T-TiSe<sub>2</sub> sample. The temporal pulse profile of the pump beam was characterized using the frequency-resolved optical gating technique. The pulse-width of the XUV pulses was estimated to a value  $< 10$  fs from the LAPE infrared–XUV cross-correlation traces of the Se 4*p* band signal of the 1T-TiSe<sub>2</sub> sample<sup>30</sup> (Fig. 2f).

# Fault lubrication during earthquakes

G. Di Toro<sup>1,2</sup>, R. Han<sup>3</sup>, T. Hirose<sup>4</sup>, N. De Paola<sup>5</sup>, S. Nielsen<sup>2</sup>, K. Mizoguchi<sup>6</sup>, F. Ferri<sup>1</sup>, M. Cocco<sup>2</sup> & T. Shimamoto<sup>7</sup>

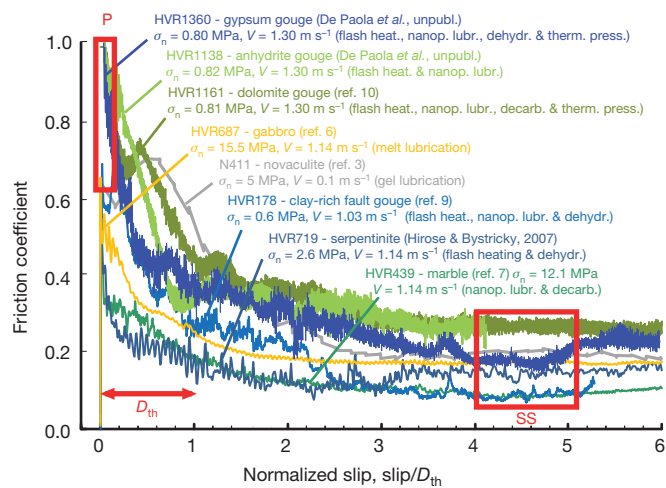
The determination of rock friction at seismic slip rates (about  $1 \text{ m s}^{-1}$ ) is of paramount importance in earthquake mechanics, as fault friction controls the stress drop, the mechanical work and the frictional heat generated during slip<sup>1</sup>. Given the difficulty in determining friction by seismological methods<sup>1</sup>, elucidating constraints are derived from experimental studies<sup>2–9</sup>. Here we review a large set of published and unpublished experiments ( $\sim 300$ ) performed in rotary shear apparatus at slip rates of  $0.1\text{--}2.6 \text{ m s}^{-1}$ . The experiments indicate a significant decrease in friction (of up to one order of magnitude), which we term fault lubrication, both for cohesive (silicate-built<sup>4–6</sup>, quartz-built<sup>3</sup> and carbonate-built<sup>7,8</sup>) rocks and non-cohesive rocks (clay-rich<sup>9</sup>, anhydrite, gypsum and dolomite<sup>10</sup> gouges) typical of crustal seismogenic sources. The available mechanical work and the associated temperature rise in the slipping zone trigger<sup>11,12</sup> a number of physicochemical processes (gelification, decarbonation and dehydration reactions, melting and so on) whose products are responsible for fault lubrication. The similarity between (1) experimental and natural fault products and (2) mechanical work measures resulting from these laboratory experiments and seismological estimates<sup>13,14</sup> suggests that it is reasonable to extrapolate experimental data to conditions typical of earthquake nucleation depths (7–15 km). It seems that faults are lubricated during earthquakes, irrespective of the fault rock composition and of the specific weakening mechanism involved.

The evolution of friction (shear stress,  $\tau$ ) during earthquakes and the dynamic friction coefficient,  $\mu$ , are key parameters in controlling seismic fault slip and radiated energy<sup>1–13</sup>. In the past 40 years, experiments performed in triaxial and biaxial apparatuses under conditions of low slip rates ( $V < 1 \text{ mm s}^{-1}$ ) and modest displacements ( $\delta < 1 \text{ cm}$ ) have shown that the friction coefficient in cohesive and non-cohesive rocks is about 0.7 (ref. 15) irrespective of the rock type (with a few exceptions that are of great relevance for the mechanics of mature faults), and that frictional instabilities of a few per cent<sup>16–19</sup>, described by rate-and-state friction laws<sup>19</sup>, are associated with earthquake initiation. Although the above results are consistent with several seismological and geophysical observations<sup>17,18</sup>, the experiments were performed at slip rates and displacements orders of magnitude smaller than those typical of earthquakes<sup>20</sup> ( $0.1\text{--}10 \text{ m s}^{-1}$  and up to 20 m, respectively). Given the low slip rates, these experiments lack a primary aspect of natural seismic slip: a large mechanical work rate (or instantaneous power density,  $\Phi(t) = \tau(t)V(t)$ ) within the slipping zone<sup>14</sup>. The work rate can be so large as to grind and mill the rock (producing particles of nanometric size, or nanopowders), trigger mechanically and thermally activated<sup>21</sup> chemical reactions, and, eventually, melt the rock<sup>22</sup>. Under these extreme deformation conditions, the fault surfaces are separated by fluids or other tribochemical products (for example melt, gel, nanopowders and decarbonation products). Work rate (not work alone) is the key parameter, as a given amount of work exchanged at a slow rate is buffered by dissipative processes and hence produces limited reactions.

In the past 15 years, the installation and exploitation of rotary shear apparatus<sup>2,3,5</sup> designed to achieve the larger slip rates and displacements typical of earthquakes produced unexpected experimental

results. Among these, the most surprising is the dramatic drop in friction (of up to 90% in most cases) at seismic slip rates, independent of the rock type and the weakening mechanism used.

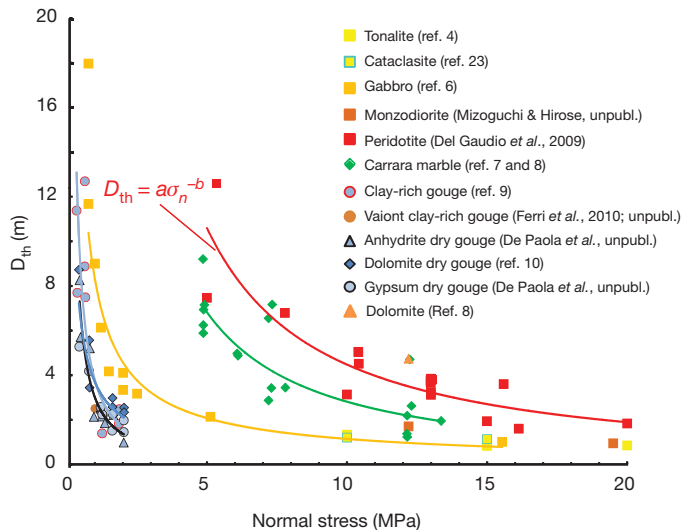
Here we report about 300 published and unpublished high-velocity rock friction experiments performed in the rotary shear apparatuses at Brown University<sup>3</sup> and at Kyoto University<sup>5</sup> (now at the Kochi Institute for Core Sample Research, JAMSTEC). These experiments were performed at room humidity with  $0.1 \text{ m s}^{-1} < V < 2.6 \text{ m s}^{-1}$ ,  $\delta > 2 \text{ m}$  and  $0.6 \text{ MPa} < \sigma_n < 20 \text{ MPa}$  (normal stress) on solid cylinders (22 and 40 mm in diameter) or hollow cylinders (15/25, 15/39, 27/39 and 40/50 mm in internal/external diameter) in the case of cohesive rock and on gouge layers confined by Teflon rings<sup>9</sup> in the case of non-cohesive rock. Figure 1 summarizes the friction coefficient as a function of normalized displacement in experiments performed at seismic slip rates for cohesive and non-cohesive rocks: both show a similar exponential decay of friction from a peak (P) to a steady-state value (SS). In all the experiments, friction decreases significantly with increasing slip. Here we introduce the thermal slip distance,  $D_{th}$ , defined as the slip distance over which the friction coefficient decays to a value  $\mu_{th} = \mu_{ss} + (\mu_p - \mu_{ss})/e$  (the experimental data are fitted by an exponential decay from a peak value,  $\mu_p$ , to a steady-state value,  $\mu_{ss}$ ).



**Figure 1 | Friction coefficient versus normalized slip.** Shear stress and slip were normalized with respect to normal stress and the thermal slip distance,  $D_{th}$ , respectively. The displacement was normalized because experiments performed with different rocks and under different normal stresses had different  $D_{th}$  values (Supplementary Information, section 4, and Fig. 2). The friction coefficient decays exponentially with slip from a peak (P) at the initiation of sliding to a steady-state (SS) value. The weakening mechanisms that we assume to be dominant are shown in parentheses (flash heat, flash heating; nanop. lubr., nanopowder lubrication; dehydr., dehydration reaction; decarb., decarbonation reaction; therm. press., thermal pressure). For all the weakening mechanisms, the friction coefficient in the steady state is  $< 0.3$ . Unpubl., unpublished experimental data. See Supplementary Information, section 2, for the reference to Hirose and Bystricky, 2007.

<sup>1</sup>Dipartimento di Geoscienze, Università di Padova, Padova 35131, Italy. <sup>2</sup>Istituto Nazionale di Geofisica e Vulcanologia, Roma 00143, Italy. <sup>3</sup>Korea Institute of Geoscience and Mineral Resources, Daejeon 305-350, South Korea. <sup>4</sup>Kochi Institute for Core Sample Research, JAMSTEC, Kochi 783-8502, Japan. <sup>5</sup>Department of Earth Sciences, University of Durham, Durham DH1 3LE, UK. <sup>6</sup>Civil Engineering Research Laboratory, Central Research Institute of Electric Power Industry, Chiba 270-1194, Japan. <sup>7</sup>Institute of Geology, China Earthquake Administration, Beijing 100029, China.





**Figure 2 | The thermal slip distance,  $D_{th}$ , versus normal stress from experiments performed with  $V = 1\text{--}1.6\text{ m s}^{-1}$ .** This plot shows that different rocks (and weakening mechanisms) have different  $D_{th}$  values and that  $D_{th}$  decreases with increasing  $\sigma_n$  (power-law dependence with a general form  $D_{th} = a\sigma_n^{-b}$ , where  $a$  and  $b$  are experimentally determined coefficients). For peridotite,  $a = 78$  and  $b = 1.24$ ; for anhydrite,  $a = 3$  and  $b = 1.13$ . By extrapolating the best-fit curves to seismogenic depths ( $\sigma_n > 200\text{ MPa}$ ), we predict that  $D_{th} \approx 7\text{ cm}$  and  $\sim 0.5\text{ cm}$  for peridotite and anhydrite, respectively. It follows that low values of friction are easily achieved during earthquakes in nature. See Supplementary Information, section 2, for references to Del Gaudio *et al.*, 2009, and Ferri *et al.*, 2010.

The thermal slip distance varies significantly between experiments performed on different rocks under the same experimental conditions (mainly the normal stress and the slip rate; Supplementary Information, section 4), but for a given rock type  $D_{th}$  seems to decrease according to a power law for increasing normal stress (Fig. 2). For instance, in the case of peridotite,  $D_{th}$  decreased from 13 m for  $\sigma_n = 5\text{ MPa}$  to 1.3 m for  $\sigma_n = 20\text{ MPa}$ . The extrapolation of these data to seismogenic depths ( $>7\text{ km}$ ), where normal stresses acting on the fault can be  $>200\text{ MPa}$ , suggests that  $D_{th}$  should decrease to less than 10 cm. Recent theoretical work<sup>6</sup> and field evidence<sup>23</sup> suggest also that  $D_{th}$  can decrease to a few centimetres in the presence of melts.

In Fig. 3, we summarize about 300 high-velocity ( $V > 0.1\text{ m s}^{-1}$ ) experiments performed on cohesive (for example novaculite, marble,

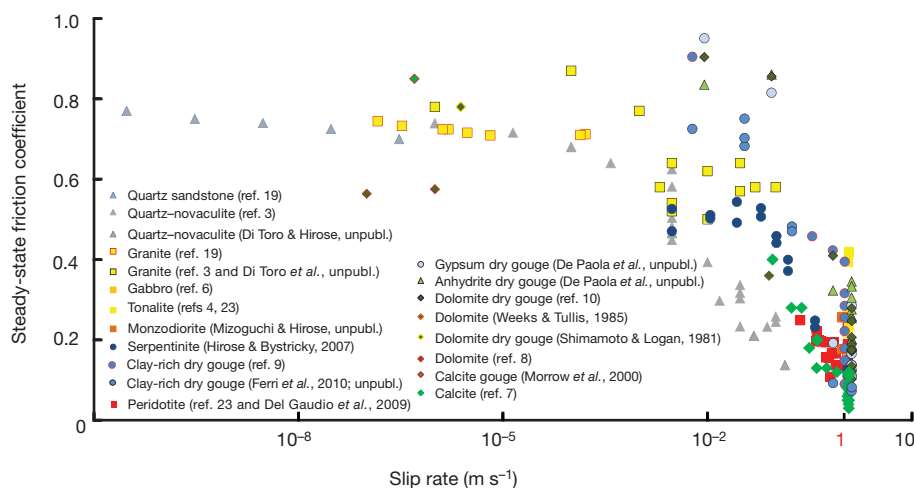
serpentinite, gabbro and tonalite) and non-cohesive rocks (anhydrite, gypsum, dolomite and clay-rich gouges) typical of crustal seismogenic sources (for data, see Supplementary Information, section 2). For each rock type there is a drastic decrease in friction that we interpret as the result of mechanically and thermally activated (that is, tribochemical), often coexisting, weakening mechanisms (see Supplementary Information, section 3, for a summary of the weakening mechanisms). For instance, in calcite-built rocks such as marbles, a significant weakening (to  $\mu = 0.03$ ) was concurrent with the emission of  $\text{CO}_2$  from the slipping zone and the production of lime and portlandite nanopowders<sup>7,8</sup>. The maximum measured average temperature was about  $900^\circ\text{C}$ , which is consistent with the thermally activated breakdown temperature for  $\text{CaCO}_3 \rightarrow \text{CaO} + \text{CO}_2$ . The measured weakening has been attributed<sup>7,8</sup> to thermal decomposition and the production of nanopowders. For melt lubrication, we considered only experiments performed at  $\sigma_n > 5\text{ MPa}$  because the nonlinear dependence of shear stress on normal stress in this case results in an overestimate of the friction coefficient at low normal stresses<sup>6</sup>. The most striking features of Fig. 3 are the velocity dependence of friction at high slip rates, the low steady-state friction at seismic slip rates and the tendency of friction to cluster around a coefficient of about 0.2–0.4 for  $V = 1\text{ m s}^{-1}$ , independently of the invoked weakening mechanism.

Figure 3 does not provide information on the mechanical work rate generated in the slipping zone, which we argue is a key parameter in controlling the rate of temperature increase ( $dT/dt \propto \Phi(t)$ : if a large amount of work is generated locally within a short time then diffusive heat loss is confined to a small rock volume, resulting in a large temperature rise in the slipping zone<sup>14</sup>) and the onset of mechanically and thermally activated weakening processes. For simplicity, we introduce an equivalent shear stress,  $\tau_e = \text{const.}$ , to determine the equivalent power density,  $\Phi_e \approx \tau_e V$ . This constant stress is a thermally balanced average of the variable stress,  $\tau(t)$ , over a time interval  $t_{th} = D_{th}/V$ . It is defined such that both  $\tau_e$  and  $\tau(t)$  yield the same temperature increase on the fault over a time  $t_{th}$ . If we assume that heat diffuses from an infinitesimally thin shear zone and use an exponential stress decay (Fig. 1), we obtain (Methods)

$$\tau_e = \tau_{ss} + \frac{\sqrt{\pi}\text{Erfi}(1)}{2e}(\tau_p - \tau_{ss})$$

Here Erfi is the imaginary error function ( $\text{Erfi}(1) = 1.65043$ ).

The approximation of an infinitesimally thin layer is based on observations of both natural and experimental seismic faults, which show that the shear band active at high slip rates generally decreases in

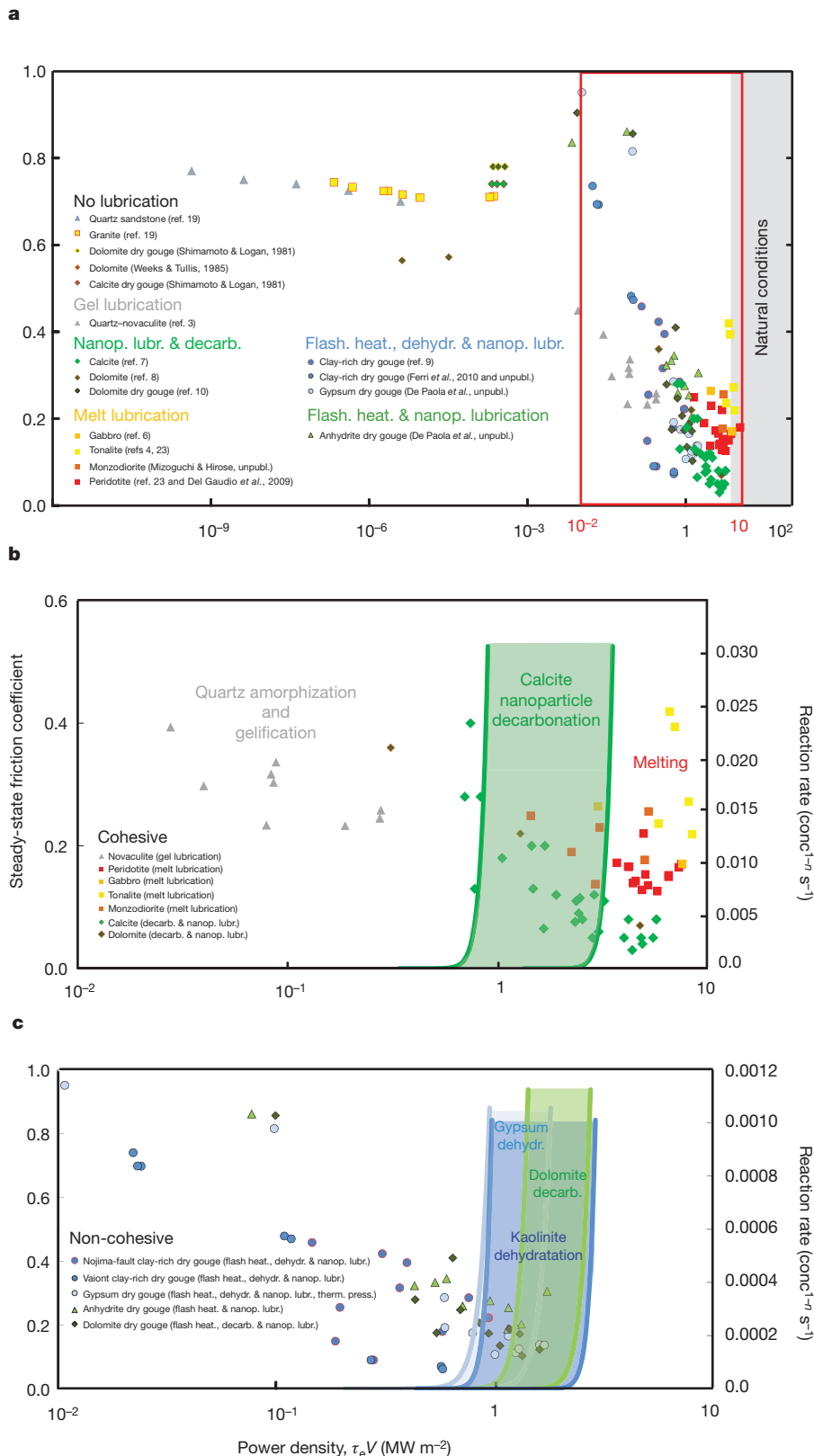


**Figure 3 | Steady-state friction coefficient versus slip rate.** There is a marked decrease in the friction coefficient when seismic slip rates ( $\sim 1\text{ m s}^{-1}$ ) are approached. Note that slip rate is plotted on a logarithmic scale here. For melt lubrication, we report only experiments performed with  $\sigma_n > 5\text{ MPa}$  (see text).

Data and extra references (Hirose and Bystricky, 2007, Ferri *et al.*, 2010, Del Gaudio *et al.*, 2009, Weeks and Tullis, 1985, Shimamoto and Logan, 1981, and Morrow *et al.*, 2000) are in Supplementary Information, section 2.

thickness to a fraction of a millimetre, even for initially thicker gouge layers<sup>7–9,21</sup>. In Fig. 4, we plot the friction coefficient as a function of the power density for cohesive (Fig. 4b) and non-cohesive (Fig. 4c) rocks (for data, see Supplementary Information, section 5). Friction data still show an evident drop approaching the power densities expected in nature (Fig. 4a, grey area), but their scaling with power density

depends on the rock type. In particular, weakening of novaculite occurs at lower power densities than does weakening of calcite- and silicate-built rocks. Figure 4a suggests the activation of different dynamic weakening mechanisms associated with different work rates and temperature increases,  $\Delta T$ , in the slipping zone. Assuming an exponential decay,  $\Delta T$  can be estimated (Supplementary Information,



**Figure 4 | Steady-state friction coefficient versus power density.** **a**, The different weakening mechanisms are activated at increasing power densities,  $\tau_e V$  (here plotted on a logarithmic scale), which also reflect the different temperatures (equation (1)) achieved in the slipping zone. The box in red is the inset shown in **b** and **c**. The grey area is for power densities expected in nature.

**b, c**, Red-boxed region of **a** for cohesive rocks (**b**) and non-cohesive rocks (**c**), showing curves of reaction speed constant,  $k$  (right-hand axis, where ‘conc’ denotes concentration, for example  $\text{mol dm}^{-3}$ , and  $n$  is the order of the reaction), versus power density. Because  $k$  varies with  $\sqrt{D_{\text{th}}/V}$  for a given power density,  $\tau_e V$  (equation (2)), for each rock type we plotted curves of reaction speed constant for the highest and lowest values of  $\sqrt{D_{\text{th}}/V}$  (Methods). In **b**, for cohesive rocks, novaculite weakening at low power densities is indicative of mechanically activated chemical reactions (amorphization and gelification of quartz<sup>2</sup>). Marble weakening is triggered at higher power densities, where thermally activated decarbonation reactions (green curves; we used  $E_A$  for decarbonation of calcite nanoparticles; see Supplementary Information, section 5) are more effective. The increase in the reaction rate constant with power density (and temperature) implies the presence of more reaction products in the slipping zone, concomitant with the boost of the weakening mechanism. In **c**, for non-cohesive rocks, dehydration and decarbonation reactions are expected to occur at higher power densities (and temperatures) than those estimated in the slipping zone, suggesting that weakening during the transient stage is dominated by mechanically activated and local, thermally activated reactions and lubrication mechanisms (nanopowder lubrication and flash weakening). Data are in Supplementary Information, section 4, and extra references (Shimamoto and Logan, 1981, Weeks and Tullis, 1985, Del Gaudio *et al.*, 2009, and Ferri *et al.*, 2010) are in Supplementary Information, section 2.

section 4) once a significant friction drop has been triggered (that is, over a slip distance  $D_{th}$ ):

$$\Delta T(t_{th} = D_{th}/V) = \frac{\tau_e V}{\rho c_p} \sqrt{\frac{D_{th}}{\pi \kappa V}} \quad (1)$$

Here  $\rho$  is the rock density,  $c_p$  is the specific heat at constant pressure and  $\kappa$  is the thermal diffusivity. Chemical reactions start once an energy barrier (the activation energy,  $E_A$ ) is overcome. The work rate,  $\tau_e V$ , and the temperature increase initiate tribochemical reactions governing dynamic fault weakening. Tribochemical reactions include mechanically and thermally activated, short-lived ( $10^{-3}$ – $10^{-6}$  s), local ( $<1 \mu\text{m}$ ) processes operating during rubbing of the sliding surfaces<sup>11,12,24–27</sup>, which result in small quantities of reaction products. Rubbing induces plastic deformation, fracturing and amorphization (mechanically activated reactions) and exposes clean, highly reactive and catalytic surfaces (which lower the  $E_A$  value of the reaction), as well as flash heating (a local and short-lived, thermally activated reaction) of the asperity contacts<sup>11,12</sup> (which increases the reaction rate). The  $E_A$  values of mechanically activated reactions are lower than the  $E_A$  values of for thermally activated reactions, and when the reaction is simultaneous with milling, the reaction rate is nearly independent of the bulk temperature<sup>12–24–26</sup> (Supplementary Information, section 5). Eventually, the bulk temperature of the slipping zone increases, direct thermal activation is more effective than mechanical excitation<sup>12</sup> and larger quantities of reaction products are produced. At this stage, the reaction rate is proportional to the reaction speed constant,  $k$  (Arrhenius-type dependence on temperature<sup>12</sup>):

$$k = A \exp\left(\frac{-E_A}{R(T_{amb} + \Delta T)}\right) \\ = A \exp\left[\frac{-E_A}{R(T_{amb} + (\tau_e V / \rho c_p) \sqrt{D_{th} / V \pi \kappa})}\right] \quad (2)$$

Here  $A$  is the pre-exponential factor,  $R$  is the gas constant and  $T_{amb}$  is room temperature.

The plot in Fig. 4b, for cohesive rocks, distinguishes weakening resulting from mechanically and local thermally activated reactions (amorphization and gelification of quartz<sup>2</sup>) from that resulting from reactions dominated by thermal activation (decarbonation of calcite<sup>7</sup>). For a power density of  $2 \text{ MW m}^{-2}$ , which corresponds to an estimated temperature increase of about  $800$ – $1,000^\circ\text{C}$  (see equation (1); Supplementary Information, section 4), in the range of measured temperatures<sup>7</sup>, we observe the limit between decarbonation and melting.

For experiments performed on non-cohesive rocks, large weakening occurs before the thermally activated, Arrhenius-type reaction (equation (2)) should dominate, suggesting the operation of other lubricating mechanisms (such as flash weakening at the asperity contacts<sup>14,28</sup> and nanopowder lubrication<sup>8,29</sup>) controlled by mechanically activated and local, thermally activated reactions<sup>11,12,24,25,27</sup>.

In summary, high-velocity experiments indicate that at seismic slip rates the friction coefficient decreases to about  $10$ – $30\%$  of its initial value (Figs 1 and 3). The extremely low friction data reported in Fig. 3 can occur in nature, implying large breakdown stress decreases (defined as the difference between the peak and the residual shear stress or, in the experiments, the steady-state shear stress) at least in some fault patches<sup>30</sup>. In fact, the thermal weakening distance,  $D_{th}$ , decreases with increasing normal stress (Fig. 2): at typical crustal stresses ( $>20 \text{ MPa}$ ),  $D_{th} < 1 \text{ m}$ , which is below the typical displacement for moderate to large earthquakes. Moreover, our experimental estimate of the breakdown work,  $W_b$ , is consistent with the seismological observations. The breakdown work, which is one of the most robust earthquake source parameters retrieved from seismological data<sup>13</sup>, is the seismological equivalent of the fracture energy<sup>14</sup>, or the energy spent per unit fault area for the advancement of rupture. Following refs 13, 14, we define  $W_b$  as the mechanical dissipation

associated with the breakdown stress drop. In the case of an exponential decay, for  $\delta \gg D_{th}$

$$W_b = \int_0^\delta (\tau_p - \tau_{ss}) \exp\left(-\frac{\delta'}{D_{th}}\right) d\delta' \approx \sigma_n (\mu_p - \mu_{ss}) D_{th} \quad (3)$$

the integral of the experimental shear stress curves. Equation (3) yields  $W_b$  values ranging between  $1$  and  $42 \text{ MJ m}^{-2}$  for most of the activated weakening mechanisms (Supplementary Information, section 4). These values are in the range of seismological  $W_b$  estimates<sup>13,14</sup> for moderate to large earthquakes ( $1 \text{ MJ m}^{-2} < W_b < 100 \text{ MJ m}^{-2}$ ). But earthquakes occur under higher normal stress than in these experiments; because  $D_{th} = a\sigma_n^{-b}$  (with  $1.13 < b < 1.24$ ; Fig. 2), from equation (3) we find that  $W_b \approx (\mu_p - \mu_{ss}) a \sigma_n^{-0.18}$ . As a consequence, we expect a decrease in the breakdown work of  $30$ – $35\%$  for an increase in  $\sigma_n$  from  $20$  to  $200 \text{ MPa}$  (that is, extrapolation of experimental data to natural conditions). Because field<sup>31</sup> and theoretical<sup>13</sup> investigations suggest that  $W_b$  is mostly converted to heat, the comparison between natural and experimental  $W_b$  values indicates that the individual processes governing dynamic fault weakening on experimental faults are similar to those governing natural faults. This is confirmed by the presence, in natural seismogenic faults, of fault products (solidified melts<sup>4,22</sup>, reaction products<sup>21</sup>, fluidized gouges<sup>9</sup> and so on) similar to those produced in the experiments<sup>4,7–9,32</sup>.

We conclude that the experimental work performed using high-velocity friction apparatus indicates that faults are lubricated when they are deformed at slip rates typical of earthquakes, independent of the rock and weakening mechanism involved. Because experiments have been performed at normal stresses ( $<25 \text{ MPa}$ ) lower than those expected in the Earth's crust, some extrapolation is required to apply these results to seismogenic faults. A more direct verification may become available when new apparatus is developed that can perform experiments under the normal stresses expected in the crust ( $>50 \text{ MPa}$ ).

## METHODS SUMMARY

The unpublished experiments cited in Figs 1–4 by De Paola *et al.*, Di Toro and Hirose, Mizoguchi and Hirose, and Ferri *et al.* were performed in the rotary shear apparatus in Kyoto University (now at the Kochi Institute for Core Sample Research, JAMSTEC, Japan). Experiments were conducted at room humidity with  $0.1 \text{ m s}^{-1} < V < 1.3 \text{ m s}^{-1}$  and  $\delta > 2 \text{ m}$  on solid cylinders ( $25 \text{ mm}$  in diameter) or hollow cylinders ( $15/25 \text{ mm}$  in internal/external diameter) in the case of cohesive rock and on  $1\text{-mm}$ -thick gouge layers sandwiched between two solid rock cylinders and confined by Teflon rings<sup>9</sup> in the case of non-cohesive rocks. In experiments performed on non-cohesive rocks, the normal stress was in the range  $0.6 \text{ MPa} < \sigma_n < 2 \text{ MPa}$ , given the low tensile strength of Teflon. For cohesive rocks, the normal stress was in the range  $5 \text{ MPa} < \sigma_n < 20 \text{ MPa}$ . The reliability of the experimental procedure and the quality of the unpublished and published experiments reported in this study are indicated by the observations that (1) the same rock had similar frictional behaviour in different rotary shear apparatuses (and also in the torsional Kolsky bar<sup>33</sup>, not discussed here; Supplementary Information, section 1); (2) different materials, such as metals, had different mechanical behaviour under similar experimental conditions (Supplementary Information, section 1); and (3) one data set (for novaculite<sup>3</sup>) covers the whole slip-rate interval (from  $1 \mu\text{m s}^{-1}$  to  $0.1 \text{ m s}^{-1}$ ) and the transition from high friction values (with results compatible with those obtained in conventional biaxial and triaxial apparatuses in the overlapping low-slip-rate range).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 29 January 2010; accepted 17 January 2011.

- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 2002).
- Goldsby, D. L. & Tullis, T. E. Low frictional strength of quartz rocks at subseismic slip rates. *Geophys. Res. Lett.* **29**, 1844 (2002).
- Di Toro, G., Goldsby, D. L. & Tullis, T. E. Friction falls towards zero in quartz rock as slip velocity approaches seismic rates. *Nature* **427**, 436–439 (2004).
- Di Toro, G. & Hirose, T., Nielsen, S., Pennacchioni, G. & Shimamoto, T. Natural and experimental evidence of melt lubrication of faults during earthquakes. *Science* **311**, 647–649 (2006).



5. Hirose, T. & Shimamoto, T. Growth of molten zone as a mechanism of slip weakening of simulated faults in gabbro during frictional melting. *J. Geophys. Res.* **110**, B05202 (2005).
6. Nielsen, S., Di Toro, G., Hirose, T. & Shimamoto, T. Frictional melt and seismic slip. *J. Geophys. Res.* **113**, B01308 (2008).
7. Han, R., Shimamoto, T., Hirose, T., Ree, J.-H. & Ando, J. Ultralow friction of carbonate faults caused by thermal decomposition. *Science* **316**, 878–881 (2007).
8. Han, R., Hirose, T. & Shimamoto, T. Strong velocity weakening and powder lubrication of simulated carbonate faults at seismic slip rates. *J. Geophys. Res.* **115**, B03412 (2010).
9. Mizoguchi, K., Hirose, T., Shimamoto, T. & Fukuyama, E. High-velocity frictional behavior and microstructure evolution of fault gouge obtained from Nojima fault, southwest Japan. *Tectonophysics* **471**, 285–296 (2009).
10. De Paola, N. *et al.* Fault lubrication and earthquake propagation in thermally unstable rocks. *Geology* **39**, 35–38 (2011).
11. Fox, P. G. Mechanically initiated chemical reactions in solids. *J. Mater. Sci.* **10**, 340–360 (1975).
12. Fisher, T. E. Tribochemistry. *Annu. Rev. Mater. Sci.* **18**, 303–323 (1988).
13. Cocco, M. & Tinti, E. Scale dependence in the dynamics of earthquake propagation: evidence from seismological and geological observations. *Earth Planet. Sci. Lett.* **273**, 123–131 (2008).
14. Rice, J. R. Heating and weakening of faults during earthquake slip. *J. Geophys. Res.* **111**, B05311 (2006).
15. Byerlee, J. D. Friction of rocks. *Pure Appl. Geophys.* **116**, 615–626 (1978).
16. Brace, W. F. & Byerlee, J. D. Stick slip as a mechanism for earthquakes. *Science* **168**, 990–992 (1966).
17. Scholz, C. H. Earthquakes and friction laws. *Nature* **391**, 37–42 (1998).
18. Marone, C. Laboratory-derived friction laws and their application to seismic faulting. *Annu. Rev. Earth Planet. Sci.* **26**, 643–696 (1998).
19. Dieterich, J. H. Modeling of rock friction 1. Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).
20. Heaton, T. H. Evidence for and implications of self healing pulses of slip in earthquake rupture. *Phys. Earth Planet. Inter.* **64**, 1–20 (1990).
21. Hirono, T. *et al.* A chemical kinetic approach to estimate dynamic shear stress during the 1999 Taiwan Chi-Chi earthquake. *Geophys. Res. Lett.* **34**, L19308 (2007).
22. Sibson, R. H. Generation of pseudotachylite by ancient seismic faulting. *Geophys. J. R. Astron. Soc.* **43**, 775–794 (1975).
23. Di Toro, G., Hirose, T., Nielsen, S. & Shimamoto, T. in *Radiated Energy and the Physics of Faulting* (eds Abercrombie, R., McGarr, A., Di Toro, G. & Kanamori, H.) 121–134 (Geophys. Monogr. Ser. 170, American Geophysical Union, 2006).
24. Heinicke, G. *Tribochemistry* (Carl-Hanser, 1984).
25. Hsu, S. M., Zhang, J. & Yin, Z. The nature and origin of tribochemistry. *Tribol. Lett.* **13**, 131–139 (2002).
26. Steinike, U. & Tkáčová, K. Mechanochemistry of solids—real structure and reactivity. *J. Mater. Synth. Process.* **8**, 197–203 (2000).
27. Baláz, P. *Mechanochemistry in Nanoscience and Minerals Engineering* (Springer, 2008).
28. Beeler, N. M., Tullis, T. E. & Goldsby, D. L. Constitutive relationships and physical basis of fault strength due to flash heating. *J. Geophys. Res.* **113**, B01401 (2008).
29. Reches, Z. & Lockner, D. A. Fault weakening and earthquake instability by powder lubrication. *Nature* **467**, 452–455 (2010).
30. Noda, H., Dunham, E. M. & Rice, J. R. Earthquake ruptures with thermal weakening and the operation of major faults at low overall stress levels. *J. Geophys. Res.* **114**, B07302 (2009).
31. Ma, K. F. *et al.* Slip zone and energetics of a large earthquake from the Taiwan Chelungpu-fault Drilling Project (TCDP). *Nature* **444**, 473–476 (2006).
32. Brantut, N., Schubnel, A., Rouzaud, J.-N., Brunet, F. & Shimamoto, T. High-velocity frictional properties of a clay bearing fault gouge and implications for earthquake mechanics. *J. Geophys. Res.* **113**, B10401 (2008).
33. Yuan, F. & Prakash, V. Use of a modified torsional Kolsky bar to study frictional slip resistance in rock-analog materials at coseismic slip rates. *Int. J. Solids Struct.* **45**, 4247–4263 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** G.D.T. thanks T. Tullis and D. Goldsby for introducing him to this topic. G.D.T. was supported by a Progetti di Eccellenza Fondazione Cassa di Risparmio di Padova e Rovigo and by European Research Council Starting Grant Project 205175.

**Author Contributions** G.D.T. wrote the paper. S.N. wrote the Methods. Original and unpublished experimental work reported in the paper: N.D.P., R.H., F.F., G.D.T., T.H. and K.M. Concept development: G.D.T., R.H., S.N., M.C., N.D.P., T.H. and T.S.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.D.T. ([giulio.ditoro@unipd.it](mailto:giulio.ditoro@unipd.it)).

## METHODS

**Estimate of the temperature in the slipping zone during the transient.** The general solution for temperature,  $T(t)$ , in a semi-infinite solid given a time-varying heat flow function,  $\Phi(t)$ , imposed at the boundary of the solid ( $z = 0$ ) is<sup>34</sup>

$$T(t) = \frac{1}{2\rho c_p \sqrt{\pi\kappa}} \int_0^t \frac{\Phi(t-\xi)}{\sqrt{\xi}} e^{-z^2/4\kappa\xi} d\xi \quad (4)$$

where the factor of two in the denominator of the prefactor accounts for bilateral heat diffusion in the wall rocks,  $\rho$  is the rock density,  $c_p$  is the specific heat capacity and  $\kappa$  is the thermal diffusivity. The experiments were performed for a constant slip rate ( $V(t) = V$ ) and a variable shear stress,  $\tau(t)$ . As a consequence, the heat flow is

$$\Phi(t) = \tau(t)V \quad (5)$$

The shear stress evolution with slip,  $x$ , during the transient of length  $D_{th}$  is approximated by an exponential decay in the form

$$\tau(x) = \tau_{ss} + (\tau_p - \tau_{ss})e^{-x/D_{th}} \quad (6)$$

where  $\tau_{ss}$  and  $\tau_p$  are the steady-state and peak shear stresses, respectively. Hence, for  $x = D_{th}$ , the thermally activated slip weakening distance over which a significant decrease in shear stress occurs,  $\tau = \tau_{ss} + (\tau_p - \tau_{ss})/e$  (Fig. 1). We remark that  $D_{th}$  corresponds to one-third of the slip weakening distance proposed in ref. 35, where a slip weakening distance  $d_c$  is introduced as the displacement at which  $\tau_p - \tau_{ss}$  decreases to 5% of its initial value. We consider  $D_{th}$  instead of  $d_c$  because at  $D_{th}$  the shear stress weakening (that is, the lubrication of the slipping zone) is already significant ( $\tau_p - \tau_{ss}$  has decreased to 36.8%, or  $1/e$ , of its initial value), indicating that the responsible chemical reactions or phase changes have been already triggered. Moreover, the mathematical description using  $1/e$  is simpler. By replacing  $x$  with  $t$  in equation (6) and using  $x = Vt$ , we obtain

$$\tau(t) = \tau_{ss} + (\tau_p - \tau_{ss})e^{-Vt/D_{th}}$$

and equation (5) results in

$$\Phi(t) = [\tau_{ss} + (\tau_p - \tau_{ss})e^{-Vt/D_{th}}]V \quad (7)$$

From equation (7), equation (4) is

$$T(t) = \frac{1}{2\rho c_p \sqrt{\pi\kappa}} \int_0^t \frac{[\tau_{ss} + (\tau_p - \tau_{ss})e^{-V(t-\xi)/D_{th}}]V e^{-z^2/4\kappa\xi}}{\sqrt{\xi}} d\xi \quad (8)$$

On solving the integral and considering the temperature at the boundary  $z = 0$ , we obtain the temperature at time  $t$ :

$$T(t) = \frac{V[2\tau_{ss}\sqrt{t} + (\tau_p - \tau_{ss})e^{-Vt/D_{th}}\sqrt{\pi D_{th}/V}\text{Erfi}(\sqrt{Vt/D_{th}})]}{2\rho c_p \sqrt{\pi\kappa}} \quad (9)$$

To determine the temperature at  $z = 0$  (border of the slipping zone) after a slip  $D_{th}$  has occurred, we may use equation (9) to show that

$$T(t = D_{th}/V) = \frac{\sqrt{D_{th}V}[2e\tau_{ss} + (\tau_p - \tau_{ss})\sqrt{\pi}\text{Erfi}(1)]}{2e\rho c_p \sqrt{\pi\kappa}} \quad (10)$$

where  $e = 2.7182$  and  $\text{Erfi}(1) = 1.65043$ . The temperature estimated from equation (10) has to be added to the initial (ambient) temperature to determine the actual temperature in the slipping zone. Equation (10) should overestimate the temperature increase after a slip distance  $D_{th}$ , as it does not include the heat losses by radiation from the sample, interaction of the heated slipping zone with air (the air cools the sample as the sample rotates up to 1,500 r.p.m.), expulsion of hot material from the slipping zone (for experiments performed in cohesive rocks) or heat exchange due to the tribochemical reactions triggered during frictional sliding (for example, the enthalpy energy of decarbonation of calcite is about  $1.8 \text{ MJ kg}^{-1}$ ). However, the samples used in the experiments are cylinders or rings of rock. It follows that the determination of slip rate,  $V$ , in equation (10) is not straightforward, as  $V$  increases with sample radius,  $r$  ( $V = \omega r$ , where  $\omega$  is the rotary speed). The slip rate in the slipping zone is obtained in terms of 'equivalent slip rate'  $V_e$  (refs 5, 36):

$$V_e = \frac{4\pi Rr}{3}$$

where  $R$  is the revolution rate of the motor. We refer to the equivalent slip rate simply as slip rate in our study. As a consequence, equation (10) estimates the average

temperature in the slipping zone, as it is calculated from the equivalent slip rate,  $V_e$ , and, for instance, the actual temperatures achieved at the sample edge might be slightly higher than those predicted from equation (10). This might explain the small underestimation of temperature predicted with equation (10) with respect to the temperatures measured at particular points of the sample during the experiments<sup>8</sup>. Lastly, the estimate from equation (10) of the temperature increase in the slipping zone implies the determination of thermal parameters (for example thermal diffusivity) that vary with temperature and with rock texture. We conclude that equation (10) yields a rough but useful, at least for the purposes of this study, estimate of the temperature increase in the slipping zone.

**Equivalent shear stress during the transient stage.** The equivalent shear stress,  $\tau_e$ , is a constant shear stress that would yield the same temperature at the boundary  $z = 0$  at time  $t_{th} = D_{th}/V$  as the approximate exponential decay observed in the experiments (Fig. 1). As a consequence, to determine  $\tau_e$  we equated the temperature achieved under constant  $\tau_e$  to the temperature under the exponential decay (equation (8)):

$$\begin{aligned} & \frac{1}{2\rho c_p \sqrt{\pi\kappa}} \int_0^t \frac{\tau_e V}{\sqrt{\xi}} d\xi \\ &= \frac{1}{2\rho c_p \sqrt{\pi\kappa}} \int_0^t \frac{[\tau_{ss} + (\tau_p - \tau_{ss})e^{-V(t-\xi)/D_{th}}]V}{\sqrt{\xi}} d\xi \end{aligned}$$

By simplifying and extracting  $\tau_e$ , we find that

$$\tau_e = \frac{\int_0^t [\tau_{ss} + (\tau_p - \tau_{ss})e^{-V(t-\xi)/D_{th}}] / \sqrt{\xi} d\xi}{\int_0^t (1/\sqrt{\xi}) d\xi}$$

Upon integration we obtain

$$\tau_e = \tau_{ss} + \frac{\sqrt{\pi}\text{Erfi}(1)}{2e}(\tau_p - \tau_{ss}) \quad (11)$$

with  $e = 2.7182$  and  $\text{Erfi}(1) = 1.65043$ . We used equation (11) in the power density plots reported in the main text (Fig. 4).

**Plot of the reaction-speed-constant curves in the friction versus power density.** In Fig. 4b, c, we plotted the reaction speed constant curves versus power density,  $\tau_e V$ .

By combining equations (10) and (11), we find that the temperature increase in the slipping zone after a slip distance  $D_{th}$  is

$$\Delta T(t = D_{th}/V) = \frac{\tau_e V}{\rho c_p \sqrt{\pi\kappa}} \sqrt{\frac{D_{th}}{V}}$$

As a consequence, the reaction speed constant is

$$\begin{aligned} k &= A \exp\left(\frac{-E_A}{R(T_{amb} + \Delta T)}\right) \\ &= A \exp\left(\frac{-E_A}{R\left(T_{amb} + (\tau_e V / \rho c_p) \sqrt{D_{th}/\pi\kappa}\right)}\right) \end{aligned} \quad (12)$$

where  $T_{amb}$  is the room temperature. Equation (12) allows us to determine the reaction speed constant for a given power density once  $D_{th}$  is known. Figure 4 includes experiments performed over a broad range of normal stresses and slip rates. As reported in Fig. 2, for a given rock  $D_{th}$  decreases with increasing normal stress and, for a given power density,  $\tau_e V$ , the reaction speed constant varies with  $\sqrt{D_{th}/V}$  (equation (12)). We therefore considered the smallest and largest  $\sqrt{D_{th}/V}$  values for a given rock type, resulting in the two reaction-speed-constant curves (for the same reaction) reported in Fig. 4b, c. The thermal properties ( $\rho$ ,  $c_p$  and  $\kappa$ ) of the different rocks (Supplementary Information, section 5) at a temperature of 300 K were considered in equation (12) and in Fig. 4b, c.

34. Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn, 76 (Clarendon, 1959).
35. Mizoguchi, K., Hirose, T., Shimamoto, T. & Fukuyama, E. Reconstruction of seismic faulting by high-velocity friction experiments: an example of the 1995 Kobe earthquake. *Geophys. Res. Lett.* **34**, L01308 (2007).
36. Shimamoto, T. & Tsutsumi, A. A new rotary-shear high-speed frictional testing machine: its basic design and scope of research [in Japanese with English abstract]. *J. Tecton. Res. Group Jpn* **39**, 65–78 (1994).

# Duplications of the neuropeptide receptor gene *VIPR2* confer significant risk for schizophrenia

Vladimir Vacic<sup>1,2</sup>, Shane McCarthy<sup>1</sup>, Dheeraj Malhotra<sup>1,3,4</sup>, Fiona Murray<sup>5,6</sup>, Hsun-Hua Chou<sup>3,4</sup>, Aine Peoples<sup>7</sup>, Vladimir Makarov<sup>8,9</sup>, Seungtae Yoon<sup>8,9</sup>, Abhishek Bhandari<sup>1,3,4</sup>, Roser Corominas<sup>4</sup>, Lilia M. Iakoucheva<sup>4</sup>, Olga Krastoshevsky<sup>10</sup>, Verena Krause<sup>10</sup>, Verónica Larach-Walters<sup>11</sup>, David K. Welsh<sup>4,12,13</sup>, David Craig<sup>14</sup>, John R. Kelsoe<sup>4,12,15</sup>, Elliot S. Gershon<sup>16</sup>, Suzanne M. Leal<sup>17</sup>, Marie Dell Aquila<sup>5,18</sup>, Derek W. Morris<sup>7</sup>, Michael Gill<sup>7</sup>, Aiden Corvin<sup>7</sup>, Paul A. Insel<sup>5,6</sup>, Jon McClellan<sup>19</sup>, Mary-Claire King<sup>20,21</sup>, Maria Karayiorgou<sup>22</sup>, Deborah L. Levy<sup>10</sup>, Lynn E. DeLisi<sup>23</sup> & Jonathan Sebat<sup>1,3,4,15,24</sup>

Rare copy number variants (CNVs) have a prominent role in the aetiology of schizophrenia and other neuropsychiatric disorders<sup>1</sup>. Substantial risk for schizophrenia is conferred by large (>500-kilobase) CNVs at several loci, including microdeletions at 1q21.1 (ref. 2), 3q29 (ref. 3), 15q13.3 (ref. 2) and 22q11.2 (ref. 4) and microduplication at 16p11.2 (ref. 5). However, these CNVs collectively account for a small fraction (2–4%) of cases, and the relevant genes and neurobiological mechanisms are not well understood. Here we performed a large two-stage genome-wide scan of rare CNVs and report the significant association of copy number gains at chromosome 7q36.3 with schizophrenia. Microduplications with variable breakpoints occurred within a 362-kilobase region and were detected in 29 of 8,290 (0.35%) patients versus 2 of 7,431 (0.03%) controls in the combined sample. All duplications overlapped or were located within 89 kilobases upstream of the vasoactive intestinal peptide receptor gene *VIPR2*. *VIPR2* transcription and cyclic-AMP signalling were significantly increased in cultured lymphocytes from patients with microduplications of 7q36.3. These findings implicate altered vasoactive

intestinal peptide signalling in the pathogenesis of schizophrenia and indicate the VPAC2 receptor as a potential target for the development of new antipsychotic drugs.

A majority of the rare CNVs that have been implicated in schizophrenia involve large (>500 kb) genomic regions where local segmental duplication architecture promotes frequent and nearly identical rearrangements by non-allelic homologous recombination (NAHR). Because of the high structural mutation rates at these loci, the strong phenotypic effects of the causal variants, and the excellent power of most array platforms to detect such large CNVs, these genomic hotspots were the first to be detected in studies of CNVs in schizophrenia. As most of the genome lacks the duplication architecture of the NAHR hotspots described earlier and because a variety of mutational mechanisms can give rise to structural rearrangements, causal variants in other regions of the genome may consist of CNVs that are individually rarer and smaller (<500 kb) than those arising at NAHR hotspots. For example, microdeletions of the gene *neurexin 1* (*NRXN1*), which are highly enriched in autism and schizophrenia<sup>6,7</sup>,

**Table 1 | Significant association of four CNV regions with schizophrenia**

| Region (hg18)             | Genes                            | Band    | Type | Primary Cases | Primary Controls | Secondary Cases | Secondary Controls | Peak OR (CI)            | Peak P-value          | Permutation P-value    |
|---------------------------|----------------------------------|---------|------|---------------|------------------|-----------------|--------------------|-------------------------|-----------------------|------------------------|
| chr22: 19786712–19795854  | <i>BCR</i>                       | 22q11.2 | Del. | 2             | 0                | 22              | 0                  | 14.21* (4.24, infinity) | $2.4 \times 10^{-6}$  | $<5.00 \times 10^{-6}$ |
| chr7: 158731401–158810016 | <i>VIPR2</i> **                  | 7q36.3  | Dup. | 2             | 0                | 18              | 1                  | 16.41 (3.11, infinity)  | $8.39 \times 10^{-5}$ | $4.00 \times 10^{-5}$  |
| chr16: 29569647–30209382  | 28 genes                         | 16p11.2 | Dup. | 4             | 0                | 18              | 1                  | 16.14 (3.06, infinity)  | 0.000097              | 0.0001                 |
| chr15: 29694064–29705665  | <i>OTUD7A</i>                    | 15q13.3 | Del. | 2             | 0                | 16              | 1                  | 14.94 (2.80, infinity)  | 0.00023               | 0.00016                |
| chr7: 158448321–158605936 | <i>VIPR2</i> , <i>BC042556</i>   | 7q36.3  | Dup. | 2             | 0                | 12              | 0                  | 8.26* (2.36, infinity)  | 0.00086               | 0.0007                 |
| chr15: 28881608–28991107  | <i>MTMR15</i>                    | 15q13.3 | Del. | 2             | 0                | 16              | 1                  | 14.94 (2.80, infinity)  | 0.00023               | 0.001                  |
| chr3: 196826549–196872080 | <i>CR597873</i> , <i>SDHALP2</i> | 3q29    | Dup. | 2             | 0                | 8               | 0                  | 5.65* (1.56, infinity)  | 0.01                  | 0.005                  |
| chr6: 162835583–162997592 | <i>PARK2</i>                     | 6q26    | Dup. | 2             | 0                | 6               | 0                  | 4.41* (1.17, infinity)  | 0.03                  | 0.044                  |

Events, ORs and exact conditional (EC) P-values listed here correspond to the peak of association. Empirical P-values for the entire target region were then computed based on permutation of case and control labels. The minimal threshold for statistical significance after Bonferroni correction for the 114 loci tested was permutation  $P < 4.4 \times 10^{-4}$ . When the number of controls in the secondary sample was 0, Haldane correction (adding 0.5 to each cell in the table) was applied in order to get a finite OR (\*). All genes overlapping with the target region are listed or the closest gene within 100 kb (\*\*). Del., deletion; Dup., duplication.

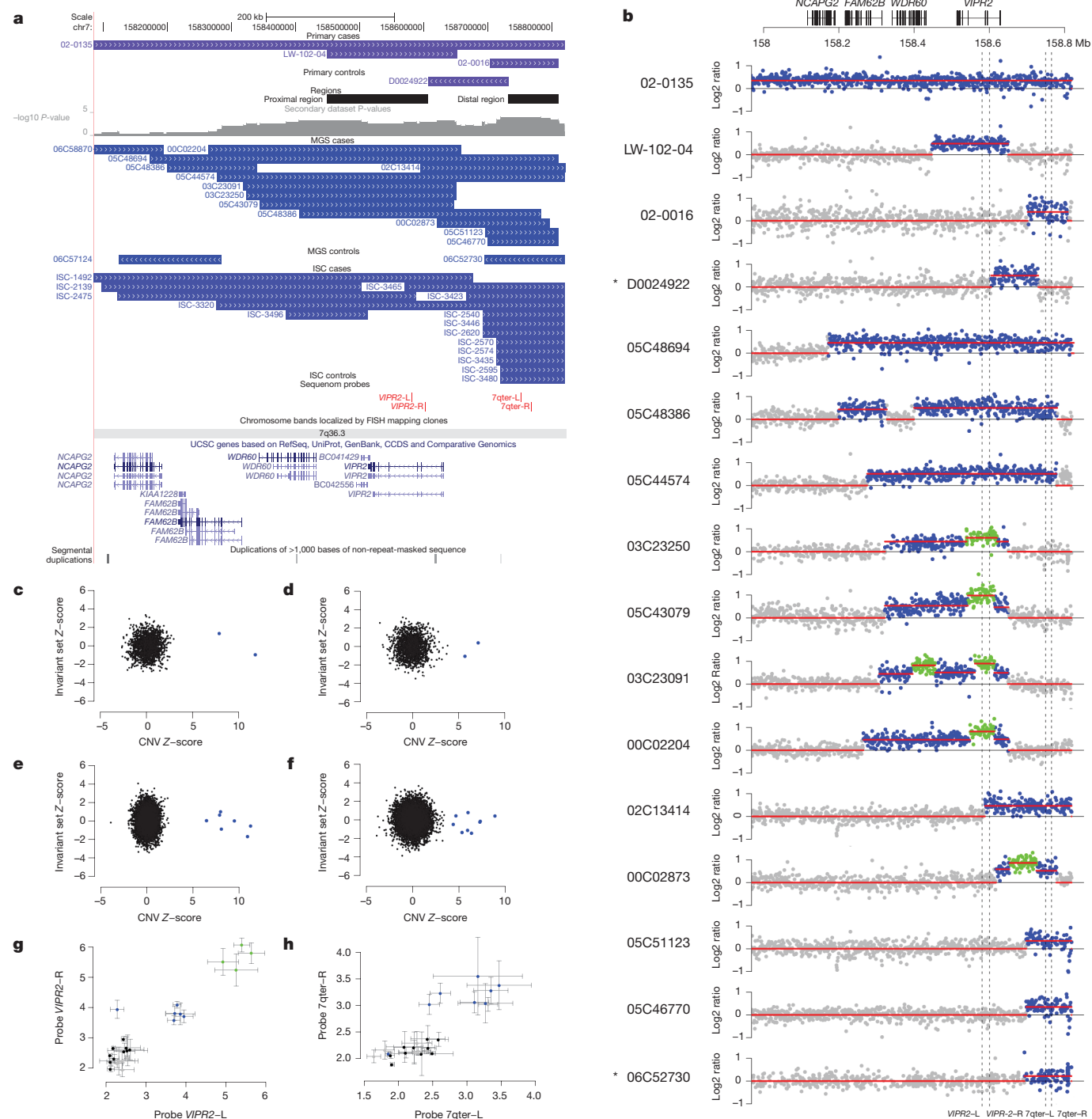
<sup>1</sup>Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA. <sup>2</sup>Department of Computer Science, Columbia University, New York, New York 10027, USA. <sup>3</sup>Beyster Center for Genomics of Psychiatric Diseases, University of California, San Diego, La Jolla, California 92037, USA. <sup>4</sup>Department of Psychiatry, University of California, San Diego, La Jolla, California 92037, USA. <sup>5</sup>Department of Medicine, University of California, San Diego, La Jolla, California 92037, USA. <sup>6</sup>Department of Pharmacology, University of California, San Diego, La Jolla, California 92037, USA. <sup>7</sup>Neuropsychiatric Genetics Research Group, Institute of Molecular Medicine and Department of Psychiatry, Trinity College Dublin, Dublin 2, Ireland. <sup>8</sup>Seaver Autism Center, Mount Sinai School of Medicine, New York, New York 10029, USA. <sup>9</sup>Department of Psychiatry, Mount Sinai School of Medicine, New York, New York 10029, USA. <sup>10</sup>McLean Hospital, Belmont, Massachusetts 02478, USA. <sup>11</sup>Faculty of Medicine, Universidad Andrés Bello, Santiago, Chile. <sup>12</sup>Veterans Affairs San Diego Healthcare System, San Diego, California 92161, USA. <sup>13</sup>Center for Chronobiology, University of California, San Diego, La Jolla, California 92037, USA. <sup>14</sup>Neurogenetics Division, Translational Genomics Research Institute, Phoenix, Arizona 85004, USA. <sup>15</sup>Institute for Genomic Medicine, University of California, San Diego, La Jolla, California 92037, USA. <sup>16</sup>Department of Psychiatry and Behavioral Neuroscience, The University of Chicago, Chicago, Illinois 60637, USA. <sup>17</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77025, USA. <sup>18</sup>Division of Medical Genetics, University of California, San Diego, La Jolla, California 92037, USA. <sup>19</sup>Department of Psychiatry, University of Washington, Seattle, Washington 98195, USA. <sup>20</sup>Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. <sup>21</sup>Department of Medicine, University of Washington, Seattle, Washington 98195, USA. <sup>22</sup>Department of Psychiatry, Columbia University, New York, New York 10027, USA. <sup>23</sup>Department of Psychiatry, Boston VA Healthcare System and Harvard Medical School, Brockton, Massachusetts 02301, USA. <sup>24</sup>Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92037, USA.



consist of overlapping deletions with non-recurrent breakpoints. *NRXN1* deletions are not flanked by segmental duplications, and may occur by different mutational mechanisms such as non-homologous end joining (NHEJ) or DNA-replication-mediated rearrangement.

To identify novel schizophrenia genes, we investigated copy number variation genome-wide using an approach that detects enrichment of multiple overlapping rare variants. Regions of interest were defined in a

primary sample of 802 patients and 742 controls as genomic segments containing CNVs in at least two cases and in no controls. This discovery step yielded 114 genomic regions of interest. In the secondary cohort of 7,488 patients and 6,689 controls, we assessed the association of these regions with schizophrenia (Supplementary Table 2). All CNVs overlapping each of the 114 regions of interest were collected, and CNV breakpoints falling within each region were used to partition the region



**Figure 1 | Detection and validation of microduplications and triplications of 7q36.3.** **a**, Map of CNVs detected in the primary and secondary cohorts from the UCSC genome browser (<http://genome.ucsc.edu>). ISC, International Schizophrenia Consortium. **b**, Plots of probe intensity ratios for 16 CNVs detected in the primary and MGS data sets. All are cases with the exception of two controls who are indicated with an asterisk. Regions with estimated copy numbers of 2, 3 and 4 are highlighted in grey, blue and green, respectively. Locations of four Sequenom validation assays are shown (dashed lines).

c–f, CNV genotypes were confirmed by MeZOD cluster plots of probe intensity ratios of the proximal and distal regions and in the primary data set (c and d, respectively) and secondary data set (e and f, respectively). g, h, Absolute copy numbers were confirmed for duplications and triplications of the proximal (*VIPR2*-L and *VIPR2*-R) and distal (7qter-L and 7qter-R) regions (b, dashed lines) by Sequenom MASSarray genotyping. Error bars represent the standard deviation of three replicates.

into a series of non-overlapping segments or bins (see Supplementary Fig. 1). Significance was tested within each bin by the exact conditional test, with ethnicity and study as covariates. The segment with the minimal *P*-value was defined as the peak of association within the region, and a permutation-based multiple testing correction scheme was applied to obtain the *P*-value for the region.

Of the 114 regions detected in the first step, four had statistically significant associations in the secondary sample after Bonferroni correction ( $\alpha = 0.05/114 = 4.4 \times 10^{-4}$ ). Table 1 lists the four regions with significant *P*-values meeting this criterion and an additional four loci with nominally significant *P*-values ( $P < 0.05$ ) in the secondary cohort. Regions with significant associations were loss of copy number at 22q11.2 ( $P < 5 \times 10^{-6}$ , odds ratio (OR) = 14.2), gain at 7q36.3 ( $P = 4.0 \times 10^{-5}$ , OR = 16.4), gain at 16p11.2 ( $P = 1.0 \times 10^{-4}$ , OR = 16.1) and loss at 15q13.3 ( $P = 1.6 \times 10^{-4}$ , OR = 14.9). No significant heterogeneity was observed for these genomic regions across studies (Breslow–Day–Tarone  $P = 0.42 - 0.83$ ).

15q13.3, 16p11.2 and 22q11.2 are well-documented loci conferring increased risk for schizophrenia<sup>2,5,8</sup>. All are hotspots for NAHR, and all alleles contributing to the association consist of large deletions with similar breakpoints. By contrast, microduplications at 7q36.3 have not been previously implicated in neuropsychiatric disorders. The 7q36.3 region harboured CNVs that overlapped but differed in size and breakpoint positions (Fig. 1a). The peak of association is located in the subtelomeric region of 7q, upstream of the gene *VIPR2*. Also, ranking fifth among the associations genome-wide was another region, 125 kb proximal to the peak at 7q36.3 ( $P = 0.0007$ ; Table 1). Combining the two 7q36.3 regions into a single 362-kb region (chromosome 7 (chr7): 158448321–158810016), duplications were detected in 29 of 8,290 (0.35%) patients and 2 of 7,431 (0.03%) controls in this study. The *P*-value for the combined region in the combined sample was  $5.7 \times 10^{-7}$  and the odds ratio (OR) (95% confidence interval (CI)) was 14.1 (3.5, 123.9). A complete list of 7q36.3 CNVs is provided in Supplementary Table 3.

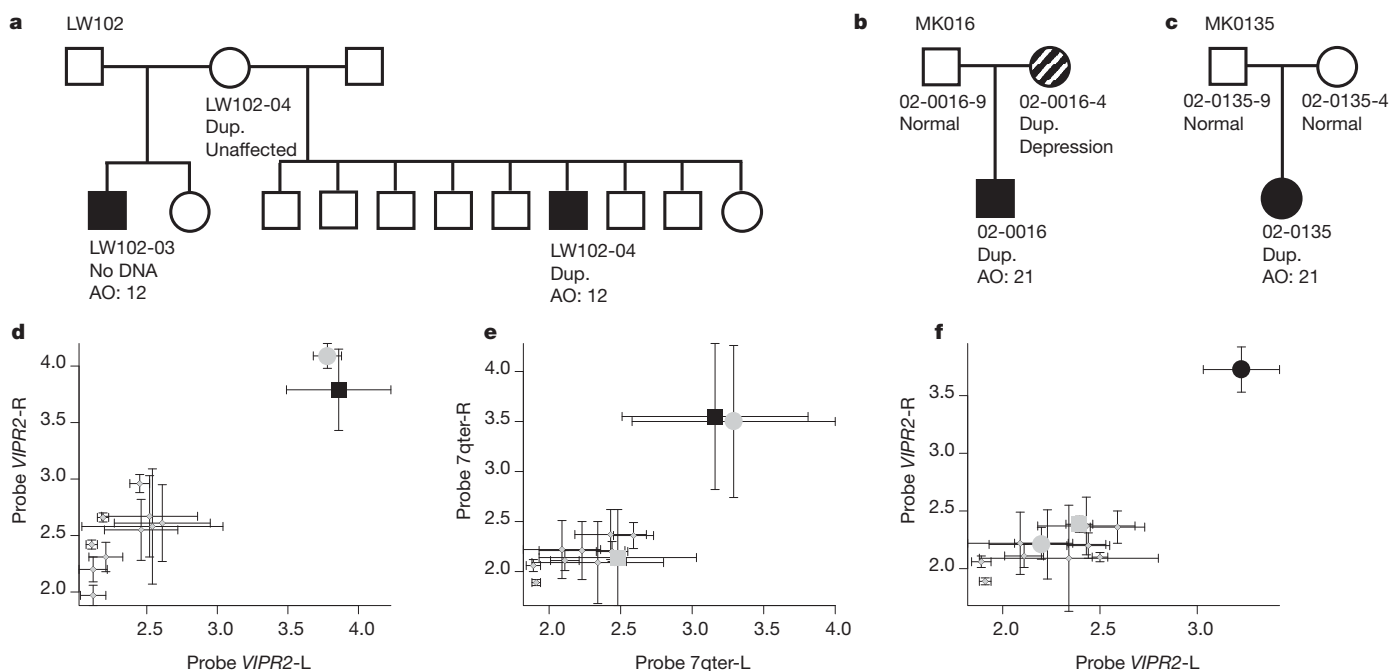
We examined sensitivity and specificity of CNV calls in the 7q36 region to determine the possibility of a spurious association (Supplementary Note). No additional duplications >100 kb were detected after

reducing the stringency of our CNV filtering criteria. Next, identical CNV calls were obtained using a more sensitive targeted CNV calling algorithm, median Z-score outlier detection (MeZOD)<sup>5</sup> (Fig. 1c–f). All but one of the duplications (control sample 06C52730) were confirmed using the Sequenom MASSarray genotyping platform with assays designed for the proximal region (Fig. 1g) and for the distal region (Fig. 1h). Validated CNVs discovered in the Molecular Genetics of Schizophrenia (MGS) study subjects were mapped at higher resolution using the NimbleGen HD2 platform, and plots of probe intensity ratios from the HD2 array are shown in Fig. 1b and Supplementary Fig. 3. In addition, tandem duplications of the *VIPR2* gene were confirmed in two patients by fluorescence *in situ* hybridization (FISH) (Supplementary Fig. 4).

Unexpectedly, manual examination of probe ratios in Fig. 1b revealed additional structural complexity within some of the 7q36.3 CNVs. Copy number profiles in four patients (03C23250, 05C43079, 03C23091 and 00C02204) indicated that there were triplications nested within duplications of the proximal region (Fig. 1b). In all four patients, a triplication overlapped with exons 3 and 4 of the gene *VIPR2*. A copy number of four was confirmed in these samples using the Sequenom MASSarray CNV assay (Fig. 1g), and results for all samples were consistent with results in Fig. 1b. *VIPR2* transcripts were amplified from messenger RNA samples from the four triplication carriers. The normal *VIPR2* transcript was detected in all samples, and we did not observe a larger product corresponding to a transcript with duplicated exons.

Inheritance of the duplication at 7q36.3 could be evaluated in three families (Fig. 2). In family 02-135, the duplication was confirmed in the proband, but not detected in either of the unaffected parents, and thus is apparently *de novo* (Fig. 2f). In family 02-016, the duplication was detected in the proband and in a mother with a diagnosis of depression (Fig. 2d). In family LW102, the duplication was detected in the proband and in an unaffected mother. The proband's mother also had a son with a diagnosis of schizophrenia (LW-102-03) from a second marriage, but DNA was not available on this individual. Clinical psychiatric reports of patients 02-016 and 02-135 are provided in the Supplementary Note.

Variable expressivity is often characteristic of pathogenic CNVs<sup>5</sup>. We evaluated the spectrum of psychiatric phenotypes associated with



**Figure 2 | Patterns of CNV inheritance in families.** a–f, Pedigree diagrams are shown for families LW102 (a), 02-016 (b) and 02-135 (c), along with the Sequenom validation for families LW102 (d), 02-016 (e) and 02-135 (f). Sequenom validation was performed on a mother and one of the affected

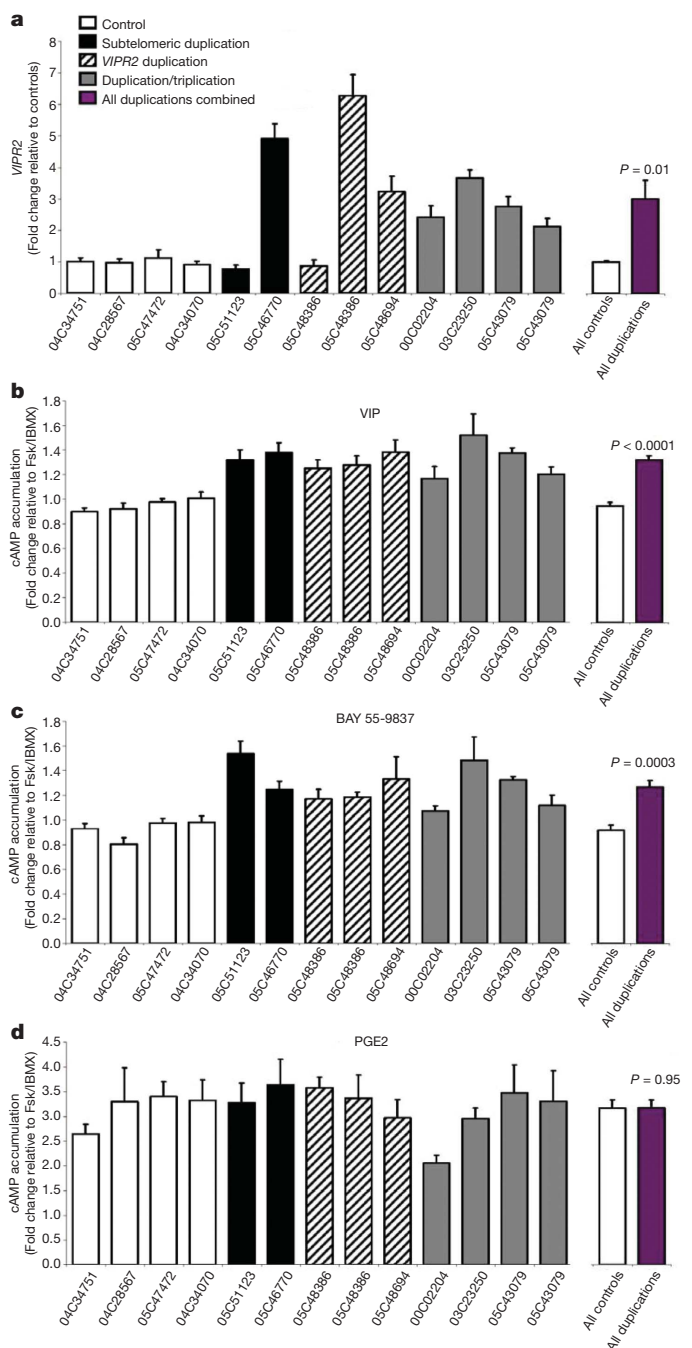
sons (a), and all three family members (b), along with 10 CEU HapMap controls. Sequenom assays confirmed that duplications (Dup.) were present in the patients and maternally inherited from LW102-2 and 02-0016-4. AO, age at onset. Error bars represent the standard deviation of three replicates.

7q36.3 duplications by screening for these events in individuals with bipolar disorder or autism spectrum disorder (ASD). Microarray data were available for 2,777 patients from the Bipolar Genome Study (BiGS), for 996 ASD patients from the Autism Genome Project Consortium (AGP), and from our unpublished analyses of 114 patients with ASD using the NimbleGen HD2 platform. Microduplications of 7q36.3 (>100 kb in size) were detected in 3 of 1,110 (0.27%) of patients with ASD; compared with the controls described earlier,  $P = 0.018$ . Microduplications at 7q36.3 were detected in 2 of 2,777 (0.07%) patients with bipolar disorder; compared with the controls,  $P = 0.23$ . These results provide preliminary evidence that the clinical phenotypes associated with 7q36.3 duplications may include paediatric neurodevelopmental disorders such as autism, but do not include bipolar disorder. Also worthy of note, larger chromosomal abnormalities involving 7q have been reported in association with neurodevelopmental disorders, including deletions of 7q36-7qter<sup>9,10</sup> and duplications of 7q35-7qter<sup>11</sup>; a 550 kb duplication of 7qter (of unknown clinical relevance) has been reported in a patient with neurofibromatosis<sup>12</sup>.

These genetic data implicate the gene *VIPR2*. All variants contributing to the observed association at 7q36.3 overlap with this gene or lie within the geneless subtelomeric region <89 kb from the transcriptional start site of *VIPR2*. *VIPR2* encodes the vasoactive intestinal peptide (VIP) receptor VPAC2, a G-protein-coupled receptor that is expressed in a variety of tissues including, in the brain, the suprachiasmatic nucleus, hippocampus, amygdala and hypothalamus<sup>13</sup>. VPAC2 binds VIP<sup>14</sup>, activates cyclic AMP (cAMP)-signalling and PKA, regulates synaptic transmission in the hippocampus<sup>15,16</sup>, and promotes the proliferation of neural progenitor cells in the dentate gyrus<sup>17</sup>. Genetic studies in mouse have established that VIP signalling has a role in learning and memory<sup>18</sup>. VPAC2 also has a role in sustaining normal circadian oscillations in the suprachiasmatic nucleus<sup>19</sup>, and *VIPR2*-null<sup>20</sup> and *VIPR2*-overexpression<sup>21</sup> mice exhibit abnormal rhythms of rest and activity.

cAMP signalling has been implicated in schizophrenia<sup>22,23</sup>. We proposed that increases in *VIPR2* transcription and VPAC2-mediated cAMP signalling would be a consequence of the microduplications at 7q36.3. We thus assessed *VIPR2* mRNA and cAMP accumulation in response to VIP and a VPAC2-selective agonist (BAY 55-9837) in lymphoblastoid cell lines from eight MGS study subjects: two with subtelomeric duplications, three with duplications of *VIPR2*, four with partial triplications, and four controls with normal copy number of the region (see Supplementary Note). *VIPR2* transcripts were present at low but measurable levels in all cell lines. *VIPR2* mRNA levels were significantly increased in duplication carriers compared with controls (Fig. 3a). Likewise, cAMP responses to VIP and BAY 55-9837 were significantly greater in lymphoblastoid lines from carriers as compared to controls (Fig. 3b). In contrast, we observed no group difference in cAMP accumulation in response to a different G-protein-coupled receptor agonist, prostaglandin E2, thus confirming that the effect of 7q36 duplications on cAMP accumulation is mediated by the VPAC2 receptor.

The expression patterns that we observe indicate that a variety of different genomic duplications can influence the transcription of *VIPR2*. The exact genetic mechanism for this is unclear. Given that some risk variants are upstream of the gene and others are complex rearrangements that could potentially disrupt the duplicate copy, our results cannot be explained simply by an increase in gene dosage. It is likely that duplications of 7q36 affect the regulation of *VIPR2*. Tandem duplication of regulatory sequences, for instance, could affect expression of the gene. Alternatively, the subtelomeric location of *VIPR2* could be relevant to the mechanism. Intrinsic regulation of telomere structure and function often affects the transcriptional regulation of adjacent genes, a phenomenon known as telomere position effect<sup>24,25</sup>. If *VIPR2* is under such epigenetic regulation, any large tandem duplication of the subtelomeric region could potentially cause the gene to escape repression. However, further studies are needed to determine the mechanism by which structural variants influence *VIPR2* expression.



**Figure 3 | Duplications and triplications of 7q36.3 result in increased *VIPR2* transcription and cAMP signalling.** **a**, Quantitative PCR results of *VIPR2* mRNA from lymphoblastoid cell lines. Two to four subjects were tested for each of four genotypes (subtelomeric duplication, *VIPR2* duplication, exon 3/4 triplication and normal diploid copy number as control). Results are expressed as the mean fold-change of the sample relative to the mean of control samples. **b, c**, cAMP accumulation was measured in the same cell lines in response to VIP (100 nM) and the VPAC2 agonist BAY 55-9837 (100 nM). Results are expressed as fold-change over forskolin (Fsk)/IBMX (3-isobutyl-1-methylxanthine, a phosphodiesterase inhibitor) alone. **d**, No significant differences were observed in cAMP response to another G-protein-coupled receptor agonist, prostaglandin E2 (PGE2, 1  $\mu$ M), demonstrating that the effects are specific to VPAC2. For subjects, error bars represent standard error of the mean computed across replicates. Differences between the groups of nine duplication carriers and four controls were tested using unpaired two-tailed *t*-test.

In light of the emerging roles of *VIPR2* in the brain, our results support the hypothesis that the pathogenesis of schizophrenia—in some patients—involves the dysregulation of cellular processes such



as adult neurogenesis and synaptic transmission and of the corresponding cognitive processes of learning and memory. Furthermore, in light of the brain expression patterns of *VIPR2* (ref. 13), our results support the involvement of certain brain regions, such as the hippocampus, amygdala and the suprachiasmatic nucleus.

The link between *VIPR2* duplications and schizophrenia may have significant implications for the development of molecular diagnostics and treatments for this disorder. Genetic testing for duplications of the 7q36 region could enable the early detection of a subtype of patients characterized by overexpression of *VIPR2*. Significant potential also exists for the development of therapeutics targeting this receptor. For instance, a selective antagonist of the VPAC<sub>2</sub> receptor could have therapeutic potential in patients who carry duplications of the *VIPR2* region. Peptide derivatives<sup>26</sup> and small molecules<sup>27</sup> have been identified that are selective VPAC<sub>2</sub> inhibitors, and these pharmacological studies offer potential leads in the development of new drugs. Although duplications of *VIPR2* account for a small percentage of patients, the rapidly growing list of rare CNVs that are implicated in schizophrenia indicates that this psychiatric disorder is, in part, a constellation of multiple rare diseases<sup>1</sup>. This knowledge, along with a growing interest in the development of drugs targeting rare disorders<sup>28</sup>, provides an avenue for the development of new treatments for schizophrenia.

## METHODS SUMMARY

**Cohort description.** Our primary cohort consisted of unrelated patients derived from family-based studies conducted by investigators at the University of Washington, McLean Hospital, Columbia University, Trinity College Dublin, New York University and Harvard Medical Schools (Supplementary Table 1). All samples were analysed by array comparative genomic hybridization (CGH) using the NimbleGen HD2 platform. The secondary cohort consisted of Affymetrix SNP Array 6.0 data from the MGS study of schizophrenia<sup>29</sup>, publicly available data from the International Schizophrenia Consortium (ISC), genotyped using Affymetrix 6.0 and 5.0 platforms<sup>3</sup>, and Affymetrix 6.0 data on an independent set of controls from the BiGS<sup>30</sup> (Supplementary Table 1).

**Intensity data processing and rare CNV calling.** With the exception of the published CNV calls from the ISC, all data were processed and analysed centrally as follows. Microarray intensity data were normalized, and CNV calls were generated using an analysis package that we developed called C-score. All CNV call sets were filtered in a consistent fashion. To minimize the differential sensitivity of the various array platforms to detect CNVs, we limited our analysis to CNVs >100 kb. This size range is readily detectable by all platforms used in this study. The same criteria have been previously applied to filter CNVs across studies<sup>3</sup>. Last, sensitivity to detect large (>100 kb) copy number polymorphisms (CNPs) was evaluated at several locations in the genome. Overall sensitivity to detect CNVs was comparable in cases and controls in both cohorts. Additional details regarding C-score, statistical methods and evaluation of CNV call sets are described in the Supplementary Note.

Received 22 October 2010; accepted 28 January 2011.

Published online 23 February 2011.

- Sebat, J., Levy, D. L. & McCarthy, S. E. Rare structural variants in schizophrenia: one disorder, multiple mutations; one mutation, multiple disorders. *Trends Genet.* **25**, 528–535 (2009).
- The International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237–241 (2008).
- Mulle, J. G. *et al.* Microdeletions of 3q29 confer high risk for schizophrenia. *Am. J. Hum. Genet.* **87**, 229–236 (2010).
- Karayorgou, M. *et al.* Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. *Proc. Natl Acad. Sci. USA* **92**, 7612–7616 (1995).
- McCarthy, S. E. *et al.* Microduplications of 16p11.2 are associated with schizophrenia. *Nature Genet.* **41**, 1223–1227 (2009).
- Szatmari, P. *et al.* Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genet.* **39**, 319–328 (2007).
- Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18**, 988–986 (2008).
- Stefansson, H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).
- Tyson, C. *et al.* Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. *Am. J. Med. Genet. A* **139A**, 173–185 (2005).
- Wu, Y. *et al.* Submicroscopic subtelomeric aberrations in Chinese patients with unexplained developmental delay/mental retardation. *BMC Med. Genet.* **11**, 72 (2010).

- Morava, E. *et al.* Small inherited terminal duplication of 7q with hydrocephalus, cleft palate, joint contractures, and severe hypotonia. *Clin. Dysmorphol.* **12**, 123–127 (2003).
- Bartsch, O. *et al.* Two independent chromosomal rearrangements, a very small (550 kb) duplication of the 7q subtelomeric region and an atypical 17q11.2 (NF1) microdeletion, in a girl with neurofibromatosis. *Cytogenet. Genome Res.* **119**, 158–164 (2007).
- Sheward, W. J., Lutz, E. M. & Harmar, A. J. The distribution of vasoactive intestinal peptide 2 receptor messenger RNA in the rat brain and pituitary gland as assessed by *in situ* hybridization. *Neuroscience* **67**, 409–418 (1995).
- Fahrenkrug, J. Transmitter role of vasoactive intestinal peptide. *Pharmacol. Toxicol.* **72**, 354–363 (1993).
- Yang, K. *et al.* Vasoactive intestinal peptide acts via multiple signal pathways to regulate hippocampal NMDA receptors and synaptic transmission. *Hippocampus* **19**, 779–789 (2009).
- Waschek, J. A. Vasoactive intestinal peptide: an important trophic factor and developmental regulator? *Dev. Neurosci.* **17**, 1–7 (1995).
- Zaben, M. *et al.* The neurotransmitter VIP expands the pool of symmetrically dividing postnatal dentate gyrus precursors via VPAC<sub>2</sub> receptors or directs them toward a neuronal fate via VPAC<sub>1</sub> receptors. *Stem Cells* **27**, 2539–2551 (2009).
- Chaudhury, D., Loh, D. H., Dragich, J. M., Hagopian, A. & Colwell, C. S. Select cognitive deficits in vasoactive intestinal peptide deficient mice. *BMC Neurosci.* **9**, 63 (2008).
- Brown, T. M., Colwell, C. S., Waschek, J. A. & Piggins, H. D. Disrupted neuronal activity rhythms in the suprachiasmatic nuclei of vasoactive intestinal polypeptide-deficient mice. *J. Neurophysiol.* **97**, 2553–2558 (2007).
- Harmar, A. J. *et al.* The VPAC<sub>2</sub> receptor is essential for circadian function in the mouse suprachiasmatic nuclei. *Cell* **109**, 497–508 (2002).
- Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
- Millar, J. K. *et al.* DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science* **310**, 1187–1191 (2005).
- Turetsky, B. I. & Moberg, P. J. An odor-specific threshold deficit implicates abnormal intracellular cyclic AMP signaling in schizophrenia. *Am. J. Psychiatry* **166**, 226–233 (2009).
- Gottschling, D. E., Aparicio, O. M., Billington, B. L. & Zakian, V. A. Position effect at S. cerevisiae telomeres: reversible repression of Pol II transcription. *Cell* **63**, 751–762 (1990).
- Koering, C. E. *et al.* Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. *EMBO Rep.* **3**, 1055–1061 (2002).
- Moreno, D. *et al.* Development of selective agonists and antagonists for the human vasoactive intestinal polypeptide VPAC<sub>2</sub> receptor. *Peptides* **21**, 1543–1549 (2000).
- Chu, A., Caldwell, J. S. & Chen, Y. A. Identification and characterization of a small molecule antagonist of human VPAC<sub>2</sub> receptor. *Mol. Pharmacol.* **77**, 95–101 (2009).
- Braun, M. M., Farag-El-Massah, S., Xu, K. & Cote, T. R. Emergence of orphan drugs in the United States: a quantitative assessment of the first 25 years. *Nature Rev. Drug Discov.* **9**, 519–522 (2010).
- Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
- Zhang, D. *et al.* Singleton deletions throughout the genome increase risk of bipolar disorder. *Mol. Psychiatry* **14**, 376–380 (2008).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This study was supported by a gift from Ted and Vada Stanley to the Cold Spring Harbor Laboratory, a gift to J.S. from the Beyster family foundation, NIH grants to J.S. (MH076431, HG04222), D.L.L. (MH071523), M.-C.K. (MH083989), P.A.I. (GM66232), F.M. (HL091061), D.K.W. (MH082945), M.K. (MH061399), L.E.D. (MH044245), grants to J.S., D.K.W., D.L.L. and M.-C.K. from NARSAD, grants to A.C., M.G., D.M. from the Wellcome Trust (072894/Z/03/Z) and Science Foundation Ireland (08INIB1916), a career development award to D.K.W. by the Veterans Administration, and grants to D.L.L. from the Sidney R. Baer, Jr Foundation and Essel Foundation. We wish to thank the Genetic Association Information Network (GAIN), Molecular Genetics of Schizophrenia (MGS) and the Bipolar Genome Study (BiGS) for providing data for this study (investigators listed in the Supplementary Note). We wish to thank B. Trask, R. Malinow and J. Gleeson for discussions.

**Author Contributions** V.V. and J.S. wrote the manuscript. V.V., S.M., D.M., H.-H.C., F.M., V.M., S.Y., S.M.L., P.A.I. and J.S. designed the analytical strategy and analysed the data. M.G., A.C., J.M., M.K., D.L.L., V.L.-W. and L.E.D. oversaw the recruitment and clinical assessment of study participants. M.D.A. performed cytogenetic analysis. A.B., A.P. and D.M. designed genotype assays and performed genotyping. H.-H.C. and R.C. performed mRNA studies. F.M. performed biochemical studies. O.K., V.K., D.W.M., V.L.-W., L.E.D. and M.K. contributed to the interpretation of clinical patient data. J.M., M.-C.K., M.K., D.L.L., L.E.D., D.C., J.R.K. and E.S.G. contributed to the interpretation of data from genetic studies. P.A.I., L.M.I. and D.K.W. contributed to the interpretation of data from functional studies. J.S. coordinated the study.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to J.S. (jsebat@ucsd.edu).

# In vitro production of functional sperm in cultured neonatal mouse testes

Takuya Sato<sup>1</sup>, Kumiko Katagiri<sup>1</sup>, Ayako Gohbara<sup>1</sup>, Kimiko Inoue<sup>2</sup>, Narumi Ogonuki<sup>2</sup>, Atsuo Ogura<sup>2</sup>, Yoshinobu Kubota<sup>1</sup> & Takehiko Ogawa<sup>1,3</sup>

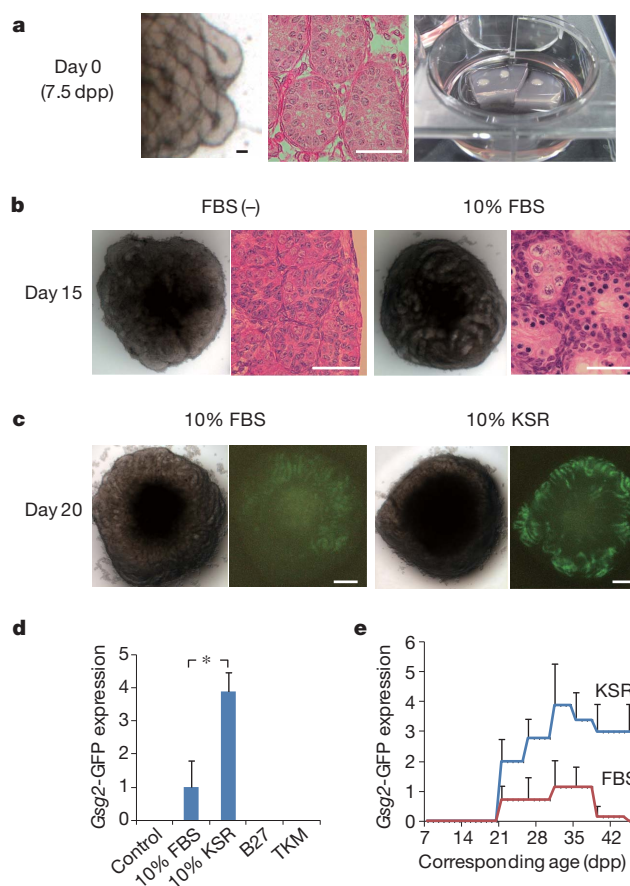
Spermatogenesis is one of the most complex and longest processes of sequential cell proliferation and differentiation in the body, taking more than a month from spermatogonial stem cells, through meiosis, to sperm formation<sup>1,2</sup>. The whole process, therefore, has never been reproduced *in vitro* in mammals<sup>3–5</sup>, nor in any other species with a very few exceptions in some particular types of fish<sup>6,7</sup>. Here we show that neonatal mouse testes which contain only gonocytes or primitive spermatogonia as germ cells can produce spermatids and sperm *in vitro* with serum-free culture media. Spermatogenesis was maintained over 2 months in tissue fragments positioned at the gas–liquid interphase. The obtained spermatids and sperm resulted in healthy and reproductively competent offspring through microinsemination. In addition, neonatal testis tissues were cryopreserved and, after thawing, showed complete spermatogenesis *in vitro*. Our organ culture method could be applicable through further refinements to a variety of mammalian species, which will serve as a platform for future clinical application as well as mechanistic understanding of spermatogenesis.

Studies on *in vitro* spermatogenesis date back to organ culture experiments about a century ago<sup>8</sup>. In 1937, it was reported that spermatogenesis proceeded up to the pachytene stage of meiosis in testis tissues of newborn mouse placed on a clot<sup>9</sup>. In the 1960s, organ culture methods had advanced and various conditions were extensively examined. However, it was not possible to promote spermatogenesis beyond the pachytene stage<sup>10,11</sup>. Thereafter, cell culture methods, instead of organ culture, were used with new concepts and devices, including immortalized germ cell lines<sup>12</sup>, the production of Sertoli cell lines as feeder cells<sup>13</sup>, bicameral chamber methods<sup>14</sup>, etc<sup>3,4</sup>. Despite such endeavours, progress has been limited, and it is still impossible to produce fertility-proven haploid cells from spermatogonial stem cells *in vitro*<sup>3–5</sup>.

At the outset of our research on *in vitro* spermatogenesis, we decided to re-evaluate organ culture methods first. According to the standard gas–liquid interphase method<sup>15</sup>, testis tissue fragments, 1–3 mm in diameter, were placed on an agarose gel half-soaked in medium (Fig. 1a). To make evaluation simple and easy, we exploited two lines of transgenic mice: *Gsg2-GFP*<sup>16</sup> (*Gsg2* is also known as *Haspin*) and *Acr-GFP*<sup>17,18</sup>, where the *GFP* gene is under control of the *Gsg2* and *Acr* promoters, respectively. These marker green fluorescent proteins (GFPs) specific for meiosis and haploid cells were extremely useful for monitoring the progress of spermatogenesis *in vitro* (Supplementary Fig. 1). Then, we devised a grading system for the extension of GFP expression to quantify the progress of spermatogenesis in each tissue (Supplementary Fig. 2).

In our previous experiments, the optimal temperature for organ culture was 34 °C, and the most effective medium was  $\alpha$ MEM (or RPMI) + 10% FBS<sup>19</sup> (Supplementary Fig. 3). Among others, FBS was indispensable to induce spermatogenesis in the organ culture experiments (Fig. 1b). Under these conditions, we found that round spermatids, haploid cells, were produced in an organ culture experiment using 7.5–10.5 days postpartum (dpp) pup mouse testis tissues<sup>19</sup>.

However, we were not able to identify any elongating spermatids or sperm. In addition, as meiosis starts around 7 dpp in mice<sup>20</sup>, the testis tissues may have included some spermatocytes from the beginning. Therefore, the study is inconclusive regarding whether or not haploid cells were produced from spermatogonia.

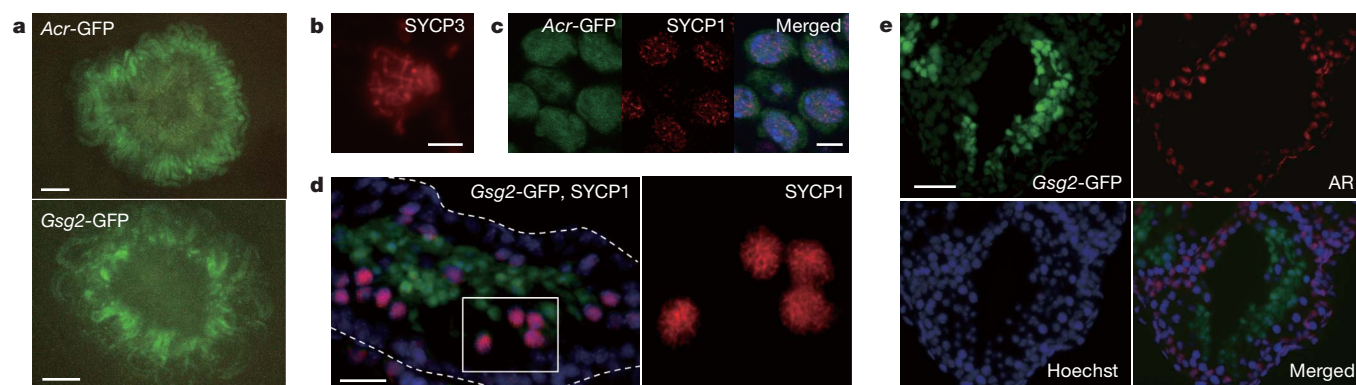


**Figure 1 | Effect of FBS and serum replacement on pup testis tissues.**

**a**, Stereomicroscopic and histological views of testis tissue fragments of 7.5 dpp pups. The tissues were placed on agarose gel stands half-soaked in the medium. **b**, Stereomicroscopic and histological views of testis tissue fragments of 7.5 dpp cultured with  $\alpha$ MEM medium without or with FBS. **c**, Testis tissue fragments of 7.5 dpp *Gsg2-GFP* transgenic mice were grown with RPMI supplemented with 10% FBS or 10% KSR. **d**, Five media, RPMI (control), RPMI + 10% FBS, RPMI + 10% KSR, RPMI + B27 and TKM, were compared on the basis of the extent of *Gsg2-GFP* expression scored using the grading scale on culture day 20 (means  $\pm$  s.d.,  $n = 6–8$ ,  $*P < 0.0001$ ). **e**, The 10% KSR induced stronger GFP expression and maintained the expression for a longer period than 10% FBS (mean  $\pm$  s.d.;  $n = 8$  and 7 for KSR and FBS, respectively). Scale bars, 50  $\mu$ m (**a**, **b**); 0.3 mm (**c**).

<sup>1</sup>Department of Urology, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan. <sup>2</sup>RIKEN, Bioresource Center, Ibaraki 305-0074, Japan. <sup>3</sup>Project Leader of Advanced Medical Research Center, Yokohama City University, Yokohama 236-0004, Japan.





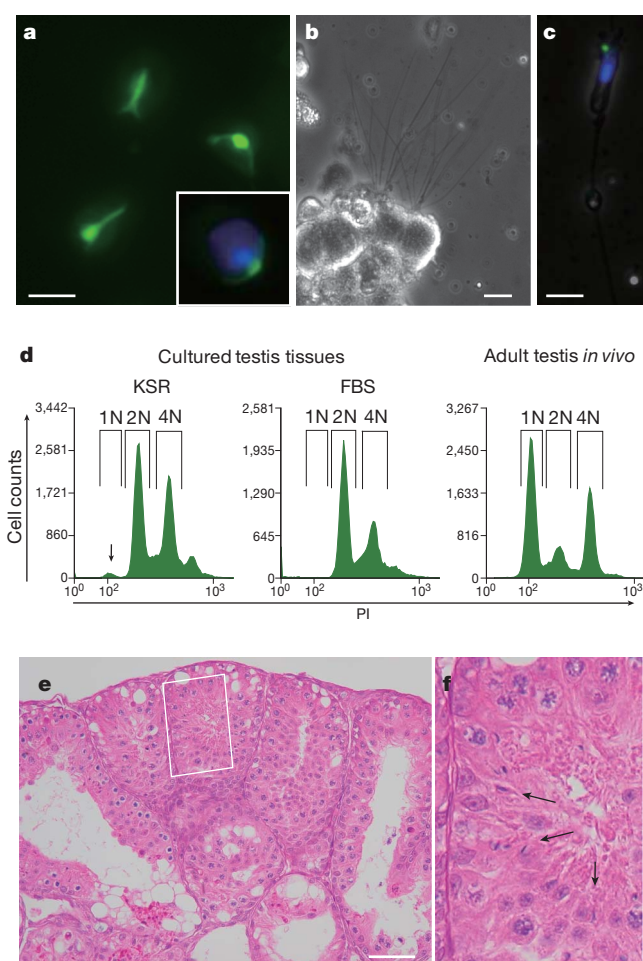
**Figure 2 | Effect of KSR on neonatal testis tissues.** **a**, 10% KSR induced the expression of both *Acr-GFP* and *Gsg2-GFP* in 2.5 and 0.5 dpp mouse testes, respectively. Pictures were taken on culture days 27 and 39, respectively. **b**, Immunostaining with anti-SYCP3 antibody. **c**, *Acr-GFP*-expressing cells (green) at the pachytene stage were also stained with SYCP1 (red). The merged picture includes Hoechst nuclear stain (blue). **d**, In a cryosection of a *Gsg2-GFP*

testis tissue, SYCP1 was demonstrated in cells (red) outside the *Gsg2-GFP*-positive cells (green). Hoechst (blue). The boxed area is enlarged in the right panel. **e**, *Gsg2-GFP*-expressing testis tissue, originating from 2.5 dpp mice and cultured for 21 days, was cryosectioned and stained with antibodies against GFP, AR, and counterstained with Hoechst dye. Scale bars, 0.5 mm (**a**); 5  $\mu$ m (**b**, **c**) and 30  $\mu$ m (**d**, **e**).

For further improvements of the culture media, we have tested different kinds of factor reported to be effective for promoting spermatogenesis or in the development of immature testes. However, none of these factors were able to promote *Acr-GFP* or *Gsg2-GFP* expression in our pilot studies (Supplementary Fig. 4). These results raised the possibility that FBS contains factors which suppress the progress of spermatogenesis, thus preventing further refinements of the culture conditions. To overcome such a possible limitation of FBS, we performed culture experiments using KSR<sup>21,22</sup> and B27 (ref. 23) as serum-replacement or serum-free TKM medium<sup>24</sup>. Whereas B27 and TKM did not induce GFP expression, KSR induced the expressions of both *Acr-GFP* and *Gsg2-GFP*. Surprisingly, the level of GFP expression induced by KSR was stronger than that induced by FBS in every experiment (Fig. 1c, d). In addition, KSR significantly extended the duration of GFP expression in culture (Fig. 1e). KSR is commonly used in culture media for embryonic stem cells to promote their proliferation while keeping them in an undifferentiated state<sup>21</sup>. However, as it has been rarely used for organ culture experiments, the present results were unexpected.

Thus, we used KSR on more immature testes of neonates, 0.5–2.5 dpp, and found the expression of both *Acr-GFP* and *Gsg2-GFP* (Fig. 2a). The effect of KSR was evident compared to that of FBS (Supplementary Fig. 5a, b). In order to confirm that such GFP expression reflected genuine meiosis, we examined the expression of meiotic marker proteins, SYCP1 and SYCP3 (ref. 5), by immunochemistry. The GFP-expressing cultured tissues were dissociated and stained with SYCP3, showing representative chromosomal spreading in some cells (Fig. 2b). When GFP and SYCP1 were costained in *Acr-GFP*-expressing cultured tissues, they were colocalized in the pachytene stage of spermatocytes (Fig. 2c). With *Gsg2-GFP* testis tissues, it was shown that SYCP1 also stained pachytene-stage spermatocytes located just on the outer side of *Gsg2-GFP*-positive cells in the seminiferous tubules, confirming that *Gsg2-GFP*-expressing cells were finishing meiosis (Fig. 2d). These results demonstrated that authentic meiosis progressed in the testis tissues cultured with KSR. The somatic cells, Sertoli and peritubular myoid cells, expressed androgen receptors (AR), a mediator of testosterone effects that is essential for spermatogenesis<sup>25</sup>, demonstrating their maturity to support spermatogenesis under the culture conditions (Fig. 2e).

Then we set experiments to find haploid cells in the cultured tissues. First, we found many spermatids in step 2–8 (refs 2, 18) in cultured samples of the *Acr-GFP* testis after the mechanical dissociation of cells into suspension in six out of the seven tissues examined, cultured for 23–50 days (Fig. 3a and Supplementary Fig. 6). In addition, we observed flagellated sperm in 5 out of the 11 samples examined, which were cultured for 27–45 days (Fig. 3b, c). These findings were also supported



**Figure 3 | Formation of spermatids.** **a**, A GFP-positive acrosomal cap was observed, indicating the presence of step 6 spermatids, in samples of dissociated testis tissues of *Acr-GFP* mouse neonates cultured for 27 days. Inset shows Hoechst-stained *Acr-GFP* spermatid. **b**, In the same sample, flagellated cells were observed. **c**, Sperm with *Acr-GFP* acrosome was stained with Hoechst. **d**, *Gsg2-GFP* mouse testis tissues, 1.5 dpp, cultured for 30 days, were subjected to flow cytometry. Tissues cultured with KSR included cells of 1N (arrow, 1.71% of all cells). The adult testis served as the control. **e**, **f**, Thin section of the tissues of 2.5 dpp *Gsg2-GFP* mice, cultured for 38 days, showed well developed seminiferous tubules showing spermatogenesis. The white box in **e** is enlarged in **f**, showing sperm formation (arrows). Scale bars, 5  $\mu$ m (**a**, **b**, **c**) and 50  $\mu$ m (**e**).



by flow cytometric analysis of dissociated cells from cultured tissues, which identified cells showing 1N ploidy as a sign of spermatids, along with 2N and 4N cells (Fig. 3d and Supplementary Fig. 7). Meanwhile, histological examinations revealed the overall phenomena occurring in the cultured tissues (Supplementary Fig. 8). In the peripheral region of each tissue piece, spermatogenesis, up to elongating spermatid formation, was observed (Fig. 3e, f). In some experiments of extended culture period, *Gsg2-GFP* expression remained at its highest level until around a corresponding age of 30–40 dpp, and then gradually decreased, but lasted beyond 70 dpp. The formation of sperm was confirmed at both 38 and 60 days of culture in a single experiment (Supplementary Fig. 9). Our organ culture system, therefore, was able to induce and maintain spermatogenesis for more than 2 months.

Finally, we tested the fertility of spermatids and sperm produced *in vitro* by microinsemination. Round spermatids retrieved from tissues, originated from 3.5 dpp testes cultured for 23 days, were used for insemination with the round spermatid injection (ROSI) technique. Sperm retrieved from the tissues, originated from 2.5 dpp testes cultured for 42 days, were used for intracytoplasmic sperm injection (ICSI) (Fig. 4a, b). Using 23 and 35 oocytes for ROSI and ICSI, respectively, 7 and 5 live offspring were delivered (Fig. 4c, Supplementary Table 1) and weaned at 3 weeks (Fig. 4d). Although these experiments were small in scale and using only a single line of mice, the efficiencies of progeny production with the *in vitro*-produced gametes were comparable to that with *in vivo*-generated counterparts<sup>26</sup>. PCR analysis of their tail tip DNA identified 4 GFP-carrying offspring out of 12, compatible with cultured testis tissues being heterozygous for *Gsg2-GFP* (Fig. 4e). Their reproductive capacity was examined by brother–sister mating, demonstrating that all four males and eight females were fertile (Supplementary Table 2).

We thought that the cryopreservation of testis tissues, if feasible, would extend our results to a variety of practical applications. Therefore, we froze neonatal testis tissues and placed them in liquid nitrogen for 4–25 days. After thawing, they were grown on agarose in

the same manner as for non-frozen samples. GFP expression was observed in all four experiments using *Acr-GFP* mice, and in two experiments using *Gsg2-GFP* mice (Supplementary Fig. 10a). In three of the four experiments using *Acr-GFP* mice, GFP-expressing acrosomes were observed in mechanically dissociated samples (Supplementary Fig. 10b). On the histological examination of cultured *Gsg2-GFP*-expressing tissues, elongated spermatids were observed in one out of the five tissues examined (Supplementary Fig. 10c). These results demonstrated that testis tissue fragments can be cryopreserved and resume full spermatogenesis later *in vitro*.

KSR was vital for the success of the present experiments. Elucidating the mechanism of its action and identifying the critical factors are important for the further refinement of our culture conditions. We have tested several factors reportedly contained in KSR<sup>22</sup>, and found that lipid-rich bovine serum albumin (AlbuMAX) was probably the most critical component regarding our present results, because the addition of AlbuMAX in place of KSR led to almost the same results (Supplementary Fig. 11a, c). In addition, it seemed that FBS does not contain factors which inhibit the progression of spermatogenesis, because medium containing both FBS and AlbuMAX was as effective in inducing spermatogenesis as that containing AlbuMAX alone (Supplementary Fig. 11b). Further studies to identify key molecule(s) in KSR and AlbuMAX are warranted.

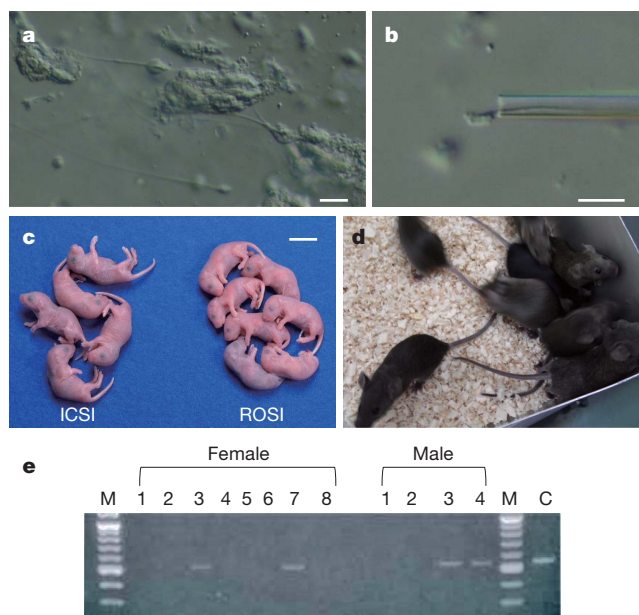
We have demonstrated that the organ culture conditions, without a circulatory system as *in vivo*, can support the complete spermatogenesis of mice. Therefore, extending the present results to a wide range of species by refinements and the individualization of culture conditions to each of them seems promising. Such progress will contribute to the elucidation of the molecular mechanisms of spermatogenesis and development of new diagnostic and therapeutic techniques for male infertility.

## METHODS SUMMARY

*Acr-GFP* and *Gsg2-GFP* transgenic mice were mated with female mice of ICR, C57BL/6, or ICR × C57BL/6F1 to produce pups. The pups were used for the culture experiments at 0.5 to 11.5 dpp. The testis tissues were placed on agarose gel half-soaked in the medium. The cultured tissues were observed every 3–7 days under a stereomicroscope equipped with an excitation light for GFP to score the extent of GFP expression in each tissue. They were also processed for histological and immunohistological examinations. For the observation of *Acr-GFP* acrosomes, cultured tissues were mechanically dissociated using needles to release cells into PBS. For cryopreservation, fragments of testis tissues were immersed in cryoprotectants for several hours or overnight at 4 °C, and then placed at –70 °C overnight before being stored in liquid nitrogen. On the initiation of culture, tissues were thawed at room temperature, soaked briefly in the medium, and then placed on agarose for culture.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 June 2010; accepted 17 January 2011.



**Figure 4 | Fertility of sperm and spermatids produced *in vitro*.** **a**, In the microinsemination experiment, elongated spermatids and spermatozoa among round spermatids were observed. **b**, Sperm were retrieved from the testis tissues of 2.5 dpp mice after culturing for 42 days. **c**, Offspring were produced by ICSI and ROSI with sperm and round spermatids, respectively. **d**, A photo of offspring at 7 weeks. **e**, Tail tip DNA analysis by PCR for *GFP*. C, DNA sample taken from a *Gsg2-GFP* mouse tail tip as a positive control; M, marker. Scale bars, 10  $\mu$ m (**a**, **b**) and 1 cm (**c**).

1. Clermont, Y. Kinetics of spermatogenesis in mammals: seminiferous epithelial cycle and spermatogonial renewal. *Physiol. Rev.* **52**, 198–236 (1972).
2. Russell, L. D., Ettlin, R. A., SinhaHikim, A. P. & Clegg, E. D. in *Histological and Histopathological Evaluation of the Testis* 1–40 (Cache River, 1990).
3. Staub, C. A century of research on mammalian male germ cell meiotic differentiation *in vitro*. *J. Androl.* **22**, 911–926 (2001).
4. Parks, J. E., Lee, D. R., Huang, S. & Kaproth, M. T. Prospects for spermatogenesis *in vitro*. *Theriogenology* **59**, 73–86 (2003).
5. La Salle, S., Sun, F. & Handel, M. A. Isolation and short-term culture of mouse spermatocytes for analysis of meiosis. *Methods Mol. Biol.* **558**, 279–297 (2009).
6. Miura, T., Yamauchi, K., Takahashi, H. & Nagahama, Y. Hormonal induction of all stages of spermatogenesis *in vitro* in the male Japanese eel (*Anguilla japonica*). *Proc. Natl Acad. Sci. USA* **88**, 5774–5778 (1991).
7. Sakai, N. Transmeiotic differentiation of zebrafish germ cells into functional sperm in culture. *Development* **129**, 3359–3365 (2002).
8. Champy, C. Quelques résultats de la méthode de culture des tissus. *Arch. Zool. Exp. Gen.* **60**, 461–500 (1920).
9. Martinovitch, P. N. Development *in vitro* of the mammalian gonad. *Nature* **139**, 413 (1937).

10. Steinberger, A., Steinberger, E. & Perloff, W. H. Mammalian testes in organ culture. *Exp. Cell Res.* **36**, 19–27 (1964).
11. Steinberger, A. & Steinberger, E. Factors affecting spermatogenesis in organ cultures of mammalian testes. *J. Reprod. Fertil., Suppl.* **2**, 117–124 (1967).
12. Feng, L. X. *et al.* Generation and *in vitro* differentiation of a spermatogonial cell line. *Science* **297**, 392–395 (2002).
13. Rassoulzadegan, M. *et al.* Transmeiotic differentiation of male germ cells in culture. *Cell* **75**, 997–1006 (1993).
14. Staub, C. *et al.* The whole meiotic process can occur *in vitro* in untransformed rat spermatogenic cells. *Exp. Cell Res.* **260**, 85–95 (2000).
15. Trowell, O. A. The culture of mature organs in a synthetic medium. *Exp. Cell Res.* **16**, 118–147 (1959).
16. Tanaka, H. *et al.* Identification and characterization of a haploid germ cell-specific nuclear protein kinase (Haspin) in spermatid nuclei and its effects on somatic cells. *J. Biol. Chem.* **274**, 17049–17057 (1999).
17. Nakanishi, T. *et al.* Real-time observation of acrosomal dispersal from mouse sperm using GFP as a marker protein. *FEBS Lett.* **449**, 277–283 (1999).
18. Ventelä, S. *et al.* Regulation of acrosome formation in mice expressing green fluorescent protein as a marker. *Tissue Cell* **32**, 501–507 (2000).
19. Gohbara, A. *et al.* *In vitro* murine spermatogenesis in an organ culture system. *Biol. Reprod.* **83**, 261–267 (2010).
20. Goetz, P., Chandley, A. C. & Speed, R. M. Morphological and temporal sequence of meiotic prophase development at puberty in the male mouse. *J. Cell Sci.* **65**, 249–263 (1984).
21. Goldsborough, M. D. *et al.* Serum-free culture of murine embryonic stem (ES) cells. *Focus* **20**, 8–12 (1998).
22. Price, P. J., Goldsborough, M. D. & Tilkins, M. L. Embryonic stem cell serum replacement. PCT/US98/00467: WO 98/30679 (1998).
23. Wachs, F. P. *et al.* High efficacy of clonal growth and expansion of adult neural stem cells. *Lab. Invest.* **83**, 949–962 (2003).
24. Tres, L. L. & Kierszenbaum, A. L. Viability of rat spermatogenic cells *in vitro* is facilitated by their coculture with Sertoli cells in serum-free hormone-supplemented medium. *Proc. Natl Acad. Sci. USA* **80**, 3377–3381 (1983).
25. Sharpe, R. M. in *Sertoli Cell Biology* (eds Skinner, M. K. & Griswold, M. D.) 199–216 (Elsevier Academic Press, 2006).
26. Ogonuki, N. *et al.* The effect on intracytoplasmic sperm injection outcome of genotype, male germ cell stage and freeze-thawing in mice. *PLoS ONE* **5**, e11062 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank S. Yoshida and G. Yoshizaki for comments and pre-submission review. We thank A. Tanaka and Y. W. Zheng for their technical help in flow cytometric analysis. T.O. is grateful to his mentor R. L. Brinster for his advice on devising the experimental strategies and for his encouragements. We would like to thank A. Steinberger and the late E. Steinberger whose work in the study of *in vitro* spermatogenesis became the ground work for our present study. This work was supported by a Grant-in-Aid for Scientific Research on Innovative Areas, “Regulatory Mechanism of Gamete Stem Cells” (#20116005); a Grant-in-Aid for Scientific Research (C) (#21592080) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan; a grant from the Yokohama Foundation for Advancement of Medical Science; and a grant for Research and Development Project II (No.S2116) of Yokohama City University, Japan (to T.O.).

**Author Contributions** T.S. performed the experiments, interpreted the results, and prepared the manuscript. K.K. performed all culture experiments. A.G. contributed to the culture experiments. K.I. and N.O. performed microinsemination experiments. A.O. performed microinsemination experiments and discussed the results. Y.K. supervised the project and discussed the results. T.O. designed and performed the experiments and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to T.O. ([ogawa@med.yokohama-cu.ac.jp](mailto:ogawa@med.yokohama-cu.ac.jp)).

## METHODS

**Animals.** *Acr-GFP* and *Gsg2-GFP* transgenic mice were provided by RIKEN BRC through the National Bio-Resource Project of MEXT, Japan. Female mice of ICR, C57BL/6 (CLEA Japan), or ICR  $\times$  C57BL/6F1 were mated with a sire of the transgenic mice to produce pups. The pups were used for the culture experiments at 0.5 to 11.5 days post-partum (dpp). All animal experiments conformed to the Guide for the Care and Use of Laboratory Animals and were approved by the Institutional Committee of Laboratory Animal Experimentation (Animal Research Center of Yokohama City University, Yokohama, Japan).

**Culture media and reagents.** The culture media used were  $\alpha$ -Minimum Essential Medium ( $\alpha$ -MEM) (Invitrogen), Roswell Park Memorial Institute 1640 (RPMI) (Invitrogen), DMEM (Dulbecco's modified Eagle's medium) (Invitrogen), and F-12 nutrient mixture (Ham) (Invitrogen). Serum and serum replacements used were fetal bovine serum (FBS) for embryonic stem cell (Gibco Invitrogen), KnockOut Serum Replacement (KSR) (Invitrogen), B-27 supplement (Gibco Invitrogen), and AlbuMAX (Invitrogen). Factors below were added to the culture media as indicated in the text. Hepatocyte growth factor (HGF) ( $5 \text{ ng ml}^{-1}$ ) (R&D Systems), Activin A ( $100 \text{ ng ml}^{-1}$ ) (Sigma-Aldrich), follicle stimulating hormone (FSH) from human pituitary ( $200 \text{ ng ml}^{-1}$ ) (Sigma-Aldrich), testosterone ( $1 \mu\text{M}$ ) (Wako Pure Chemical Industries), recombinant human BMP-4 ( $20 \text{ ng ml}^{-1}$ ) (R&D Systems), recombinant human BMP-7 ( $20 \text{ ng ml}^{-1}$ ) (R&D Systems), and bovine pituitary extract ( $50 \mu\text{g ml}^{-1}$ ) (Invitrogen).

**Culture method.** The testes of the neonatal or pup mice were decapsulated and gently separated by forceps into 1 to 8 pieces of 1–3 mm in diameter. The tissue fragments were then positioned on stands made of agarose gel placed in culture plate wells. To make the agarose gel stand, agarose-1 (Dojindo Molecular Technologies) was heated to dissolve it in distilled water (1.5% (w/v)) and then poured into a 10-cm dish. After cooling, the gels were cut into hexahedrons of about  $10 \times 10 \times 5 \text{ mm}$  in size. They were then soaked in the culture medium for more than 24 h to replace the water in them with the medium. Three to four pieces of the agarose gels were placed in the wells of a 6-well plate (Sumitomo Bakelite). Each gel was loaded with one to three testis tissue fragments. The amount of medium was adjusted so it would come up to half to four fifth of the height of the agarose gel. Medium change was performed once a week. The culture incubator was supplied with 5% carbon dioxide in air and maintained at  $34^\circ\text{C}$ .

**Observations.** The cultured tissues were observed every 3 to 7 days under a stereomicroscope equipped with an excitation light for GFP (Olympus SZX12; Olympus) to score the level of GFP expression of the tissues. For histological examination, the specimens were fixed with Bouin's fixative and embedded in paraffin. One section showing the largest cut surface was made for each specimen and stained with haematoxylin and eosin (H&E). For immunofluorescence staining, tissues fixed with 4% paraformaldehyde in PBS were cryo-embedded in OCT compound (Sakura Finetechnical) and cut into  $7\text{-}\mu\text{m}$ -thick sections. The first antibodies to be used were rabbit anti-androgen receptor (AR) antibody (1:500; Santa Cruz Biotechnology), rabbit anti-SYCP1 antibody (1:600; Novus Biologicals), and rabbit anti-GFP Alexa Fluor 488 conjugate (1:50; Invitrogen). Alexa Fluor 546-conjugated goat anti-rabbit IgG (1:400; Invitrogen) was used as a second antibody for anti-AR and anti-SYCP1 antibodies. Nuclei were counterstained with Hoechst

33342 dye. Specimens were observed with a microscope (Nikon Eclipse TE200) or confocal laser microscope (Olympus FV-1000D). For the detection of SYCP3, cultured tissues were mechanically dissociated with fine forceps, then fixed with 4% paraformaldehyde in PBS, and stained with rabbit anti-SYCP3 antibody (1:400; Abcam) followed by Alexa Fluor 546-conjugated goat anti-rabbit IgG as a second antibody. For the observation of *Acr-GFP* acrosomes, cultured tissues were mechanically dissociated using needles to release cells into PBS. The cell suspension was observed with a microscope under GFP excitation light.

**Flow cytometric analysis.** The cultured testis tissue fragments were treated with  $2 \text{ mg ml}^{-1}$  of collagenase for 15 min, followed by 0.25% trypsin/1 mM EDTA digestion for 10 min at  $37^\circ\text{C}$  to dissociate cells. After passing through a cell strainer with a  $40 \mu\text{m}$  pore size (Becton Dickinson), the cells were suspended in PBS containing 3% fetal bovine serum (FBS) and subjected to flow cytometry to analyse GFP-expressing cells using a MoFlo sorter (Dako Cytomation). For DNA ploidy, the singly-dissociated cells were fixed in 1% paraformaldehyde at  $4^\circ\text{C}$  for 15 min, followed by 70% ethanol at  $-25^\circ\text{C}$  for 12–24 h, and re-suspended and incubated in staining solution (0.1% Triton X-100 in PBS,  $0.2 \text{ mg ml}^{-1}$  RNase A,  $0.02 \text{ mg ml}^{-1}$  propidium iodide) at  $37^\circ\text{C}$  for 15 min. The flow cytometric analysis was also performed with the MoFlo sorter.

**Microinsemination.** The cultured testes tissues were dissected out under a stereomicroscope. Round spermatids or spermatozoa were collected and injected into the ooplasm of wild-type matured oocytes of B6D2F1 using a Piezo-driven micromanipulator. For fertilization with round spermatids, oocytes were then activated by treatment with  $\text{SrCl}_2$  in the presence of cytochalasin B to resume meiosis. After the formation of two female pronuclei, one was removed with a micro-pipette<sup>27</sup>. Fertilized oocytes were cultured for 24 h, and two-cell embryos were transferred into the oviducts of pseudopregnant ICR females. Live fetuses retrieved on day 19.5 were raised by lactating foster ICR dams.

**PCR analysis.** Genomic DNA was extracted from the mouse tail with a DNeasy Tissue kit (Qiagen). The DNA samples (10 ng) were added to a 20- $\mu\text{l}$  reaction mixture containing  $0.25 \mu\text{M}$  of each enhanced green fluorescent protein (EGFP)-specific primer and Premix Ex Taq (Takara Bio). EGFP-specific primers were 5'-TACGGCAAGCTGACCCTGAA-3' and 5'-TGTGATCGCGCTTCTCGTTG-3'. The reaction profile was 30 cycles of denaturation at  $95^\circ\text{C}$  for 60 s, annealing at  $60^\circ\text{C}$  for 30 s, and extension at  $72^\circ\text{C}$  for 60 s.

**Cryopreservation of testis tissues.** The fragments of testis tissues, prepared in exactly the same way as for culture, were immersed in TC-protector cell-freezing medium (BUF050, AbD Serotec) for several hours to overnight, and then frozen at  $-70^\circ\text{C}$  overnight before being placed in liquid nitrogen. The tissues were stored in liquid nitrogen for 4–25 days. On the initiation of culture, cryotubes were placed at room temperature to thaw the cryoprotectant solution and tissues were taken out to be soaked briefly in the culture medium to remove the cryoprotectants. Then, they were placed on agarose for culturing.

**Statistical analysis.** One-way analysis of variance (ANOVA) was used to compare differences between groups. Values with  $P < 0.05$  were considered significant.

27. Ogura, A., Inoue, K. & Matsuda, J. Mouse spermatid nuclei can support full term development after premature chromosome condensation within mature oocytes. *Hum. Reprod.* **14**, 1294–1298 (1999).



# Fat cells reactivate quiescent neuroblasts via TOR and glial insulin relays in *Drosophila*

Rita Sousa-Nunes<sup>1</sup>, Lih Ling Yee<sup>1</sup> & Alex P. Gould<sup>1</sup>

Many stem, progenitor and cancer cells undergo periods of mitotic quiescence from which they can be reactivated<sup>1–5</sup>. The signals triggering entry into and exit from this reversible dormant state are not well understood. In the developing *Drosophila* central nervous system, multipotent self-renewing progenitors called neuroblasts<sup>6–9</sup> undergo quiescence in a stereotypical spatiotemporal pattern<sup>10</sup>. Entry into quiescence is regulated by Hox proteins and an internal neuroblast timer<sup>11–13</sup>. Exit from quiescence (reactivation) is subject to a nutritional checkpoint requiring dietary amino acids<sup>14</sup>. Organ co-cultures also implicate an unidentified signal from an adipose/hepatic-like tissue called the fat body<sup>14</sup>. Here we provide *in vivo* evidence that Slimfast amino-acid sensing and Target of rapamycin (TOR) signalling<sup>15</sup> activate a fat-body-derived signal (FDS) required for neuroblast reactivation. Downstream of this signal, Insulin-like receptor signalling and the Phosphatidylinositol 3-kinase (PI3K)/TOR network are required in neuroblasts for exit from quiescence. We demonstrate that nutritionally regulated glial cells provide the source of Insulin-like peptides (ILPs) relevant for timely neuroblast reactivation but not for overall larval growth. Conversely, ILPs secreted into the haemolymph by median neurosecretory cells systemically control organismal size<sup>16–18</sup> but do not reactivate neuroblasts. *Drosophila* thus contains two segregated ILP pools, one regulating proliferation within the central nervous system and the other controlling tissue growth systemically. Our findings support a model in which amino acids trigger the cell cycle re-entry of neural progenitors via a fat-body–glia–neuroblasts relay. This mechanism indicates that dietary nutrients and remote organs, as well as local niches, are key regulators of transitions in stem-cell behaviour.

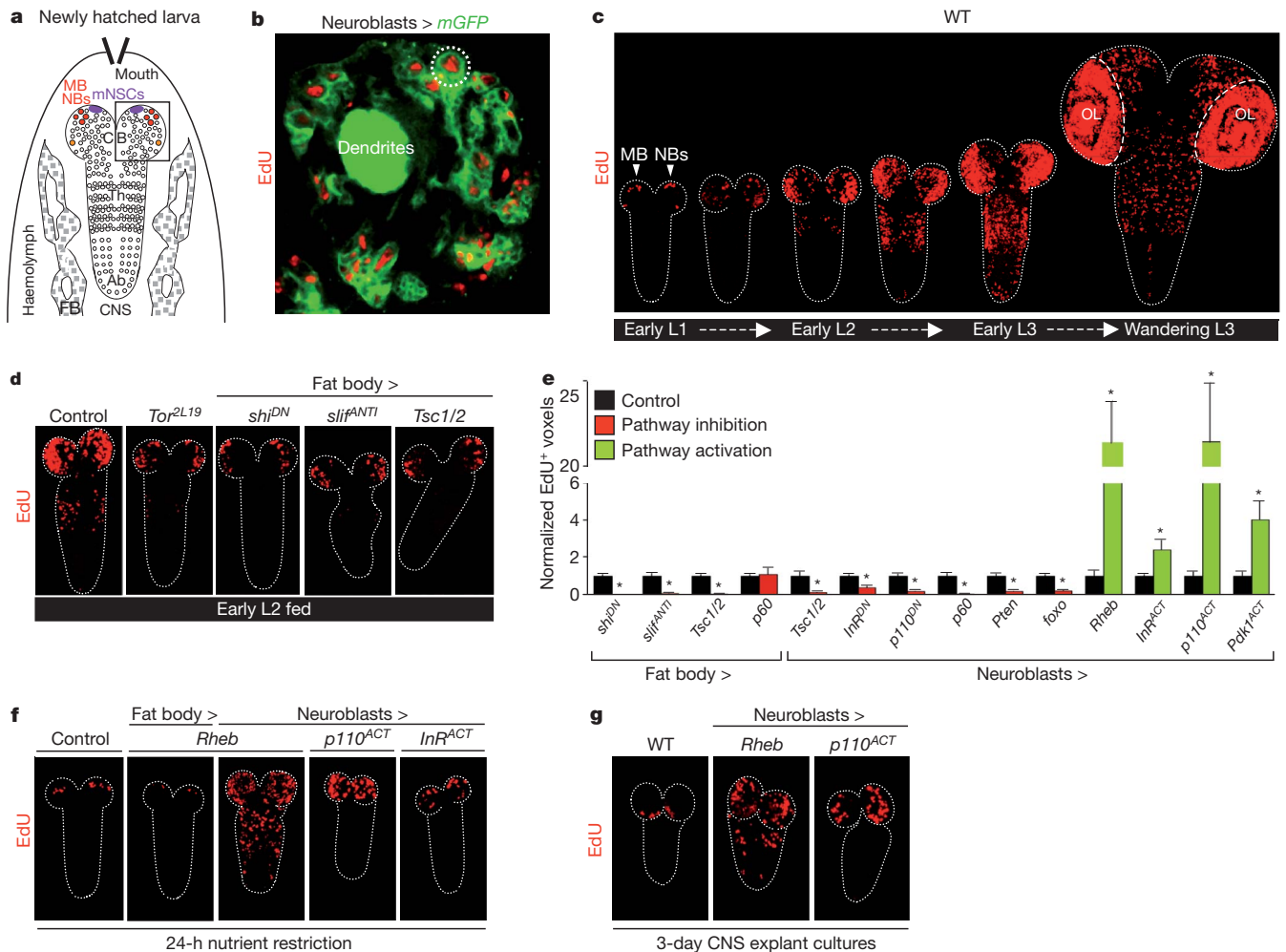
In fed larvae, *Drosophila* neuroblasts (Fig. 1a) exit quiescence from the late first instar (L1) stage onwards. This reactivation involves cell enlargement and entry into S phase, monitored in this study using the thymidine analogue 5-ethynyl-2'-deoxyuridine (EdU). Consistent with a previous study<sup>10</sup>, we observed that reactivated neuroblast lineages (neuroblasts and their progeny; Fig. 1b) reproducibly incorporated EdU in a characteristic spatiotemporal sequence: central brain → thoracic → abdominal neuromeres (Fig. 1c and Supplementary Fig. 1). Mushroom-body neuroblasts and one ventrolateral neuroblast, however, are known not to undergo quiescence and to continue dividing for several days in the absence of dietary amino acids<sup>14</sup> (Fig. 1a, c, f). This indicates that dietary amino acids are more than mere 'fuel', providing a specific signal that reactivates neuroblasts. However, explanted central nervous systems (CNSs) incubated with amino acids do not undergo neuroblast reactivation unless co-cultured with fat bodies from larvae raised on a diet containing amino acids<sup>14</sup>. We therefore tested the *in vivo* requirement for a fat-body-derived signal (FDS) in neuroblast reactivation by blocking vesicular trafficking and thus signalling from this organ using a dominant-negative Shibire dynamin (SHI<sup>DN</sup>). This strongly reduced neuroblast EdU incorporation, indicating that exit from quiescence *in vivo* requires an FDS (Fig. 1d, e). One candidate we tested was *Ilp6*, known to be expressed by the fat body<sup>19,20</sup>, but neither fat-body-specific overexpression nor RNA interference of this gene significantly affected neuroblast reactivation (Supplementary

Table 1 and data not shown). Fat-body cells are known to sense amino acids via the cationic amino-acid transporter Slimfast (SLIF), which activates the TOR signalling pathway, in turn leading to the production of a systemic growth signal<sup>15,21</sup>. We found that fat-body-specific overexpression of the TOR activator Ras homologue enriched in brain (RHEB), or of an activated form of the p110 PI3K catalytic subunit, or of the p60 adaptor subunit, had no significant effect on neuroblast reactivation in fed animals or in larvae raised on a nutrient-restricted diet lacking amino acids (Fig. 1e, f and data not shown). In contrast, global inactivation of *Tor*, fat-body-specific *Slif* knockdown or fat-body-specific expression of the TOR inhibitors Tuberous sclerosis complex 1 and 2 (*Tsc1/2*) all strongly reduced neuroblasts from exiting quiescence (Fig. 1d, e). Together, these results show that a SLIF/TOR-dependent FDS is required for neuroblasts to exit quiescence and that this may be equivalent to the FDS known to regulate larval growth.

Next we investigated the signalling pathways essential within neuroblasts for their reactivation. Nutrient-dependent growth is regulated in many species by the interconnected TOR and PI3K pathways<sup>22–24</sup> (Supplementary Fig. 2). In fed larvae, we found that neuroblast inactivation of TOR signalling (by overexpression of *TSC1/2*), or PI3K signalling (by overexpression of p60, the Phosphatase and tensin homologue PTEN, the Forkhead box subgroup O transcription factor FOXO or dominant-negative p110), all inhibited reactivation (Fig. 1e). Conversely, stimulation of neuroblast TOR signalling (by overexpression of RHEB) or PI3K signalling (by overexpression of activated p110 or Phosphoinositide-dependent kinase 1 (PDK1)) triggered precocious exit from quiescence (Fig. 1e). RHEB overexpression had a particularly early effect, preventing some neuroblasts from undergoing quiescence even in newly hatched larvae (Supplementary Fig. 3). Hence, TOR/PI3K signalling in neuroblasts is required to trigger their timely exit from quiescence. Importantly, neuroblast overexpression of RHEB or activated p110 in nutrient-restricted larvae, which lack FDS activity<sup>14</sup>, was sufficient to bypass the block to neuroblast reactivation (Fig. 1f). Notably, both genetic manipulations were even sufficient to reactivate neuroblasts in explanted CNSs, cultured without fat body or any other tissue (Fig. 1g). Together with the previous results this indicates that neuroblast TOR/PI3K signalling lies downstream of the amino-acid-dependent FDS during exit from quiescence.

To identify the mechanism bridging the FDS with neuroblast TOR/PI3K signalling, we tested the role of the Insulin-like receptor (InR) in neuroblasts (Supplementary Fig. 2). Importantly, a dominant-negative InR inhibited neuroblast reactivation, whereas an activated form stimulated premature exit from quiescence (Fig. 1e). Furthermore, InR activation was sufficient to bypass the nutrient restriction block to neuroblast reactivation (Fig. 1f). This indicates that at least one of the potential InR ligands, the seven ILPs, may be the neuroblast reactivating signal(s). By testing various combinations of targeted *Ilp* null alleles<sup>25</sup> and genomic *Ilp* deficiencies<sup>25,26</sup>, we found that neuroblast reactivation was moderately delayed in larvae deficient for both *Ilp2* and *Ilp3* (*Df(3L)Ilp2–3*) or lacking *Ilp6* activity (Fig. 2a). Stronger delays, as severe as those observed in *InR*<sup>3f</sup> mutants, were observed in larvae simultaneously lacking the activities of *Ilp2*, 3 and 5 (*Df(3L)Ilp2–3, Ilp5*) or *Ilp1–5* (*Df(3L)Ilp1–5*)

<sup>1</sup>Division of Developmental Neurobiology, Medical Research Council National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK.



**Figure 1 | TOR/PI3K signalling in fat body and neuroblasts regulates reactivation.** **a**, Diagram depicting larval fat body (FB) and CNS with central brain (CB), thoracic (Th) and abdominal (Ab) neuromeres, mNSCs, mushroom-body neuroblasts (MB NBs) and other neuroblasts (circles) indicated. **b**, Brain lobe (inset in Fig. 1a), showing EdU incorporation in postembryonic neuroblasts (large cells; for example, dotted circle) and their progeny (smaller cells), labelled with *nab-GAL4* driving membrane GFP

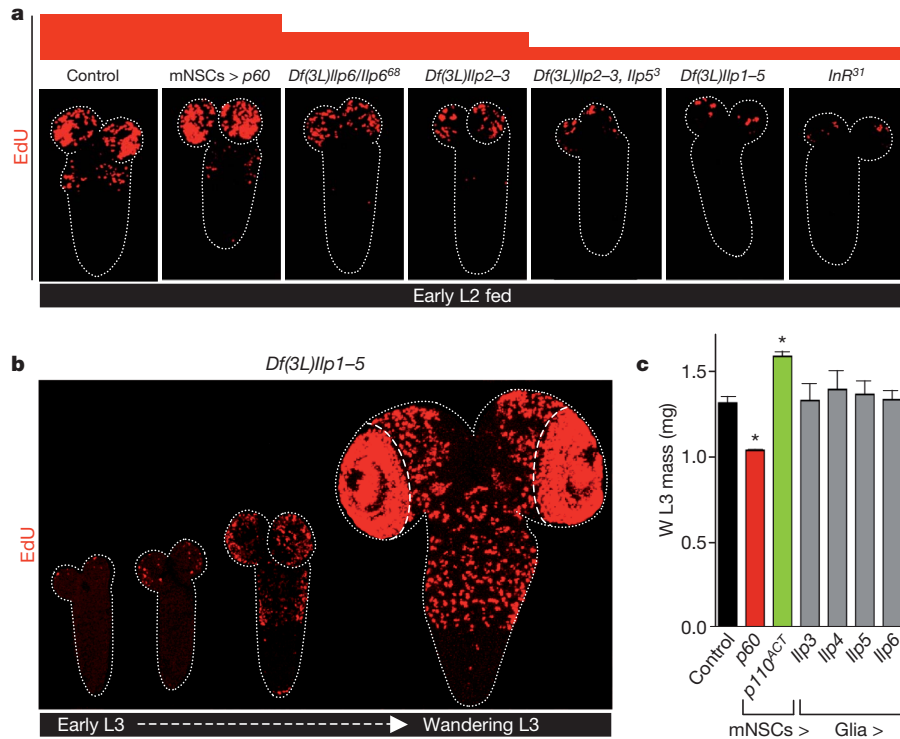
(Neuroblasts > *mGFP*). **c**, EdU incorporation time course from first-instar (L1) to third-instar (L3) larval stages in the wild-type (WT) CNS. OL, optic lobe. **d**, **f**, **g**, EdU-labelled CNSs from larvae expressing TOR/PI3K components driven by *Cg-GAL4* (Fat body >) or *nab-GAL4* (Neuroblasts >). **e**, Histograms of EdU<sup>+</sup> voxels from thoracic CNSs of fed larvae, normalized to controls. In this and all subsequent figures, error bars are s.e.m.; \**P* < 0.05. See text, Methods and Supplementary Fig. 2 for details of molecules expressed.

(Fig. 2a). Despite the developmental delay in *Df(3L)Ilp1-5* homozygotes<sup>25,26</sup>, neuroblast reactivation eventually begins in the normal spatial pattern—albeit heterochronically—in larvae with L3 morphology (Fig. 2b, compare timeline with Fig. 1c). Together, the genetic analysis shows that *Ilp2*, 3, 5 and 6 regulate the timing but not the spatial pattern of neuroblast exit from quiescence. However, as removal of some ILPs can induce compensatory regulation of others<sup>25</sup>, the relative importance of each cannot be assessed from loss-of-function studies alone.

Brain median neurosecretory cells (mNSCs) (Fig. 1a) are an important source of ILPs, secreted into the haemolymph in an FDS-dependent manner to regulate larval growth<sup>16–18,21</sup>. They express *Ilp1*, 2, 3 and 5, although not all during the same development stages<sup>16–18</sup>. However, we found that none of the seven ILPs could reactivate neuroblasts during nutrient restriction when overexpressed in mNSCs (Supplementary Table 2). Similarly, increasing mNSC secretion using the NaChBac sodium channel<sup>21</sup> or altering mNSC size using PI3K inhibitors/activators, which in turn alters body growth, did not significantly affect neuroblast reactivation under fed conditions (Fig. 2a, c, Supplementary Fig. 1b and L. Y. Cheng and colleagues, manuscript submitted). Surprisingly, therefore, mNSCs are not the relevant ILP source for neuroblast reactivation. Nonetheless, *Ilp3* and *Ilp6* messenger RNAs were detected in the CNS cortex, at the early L2 stage, in a domain

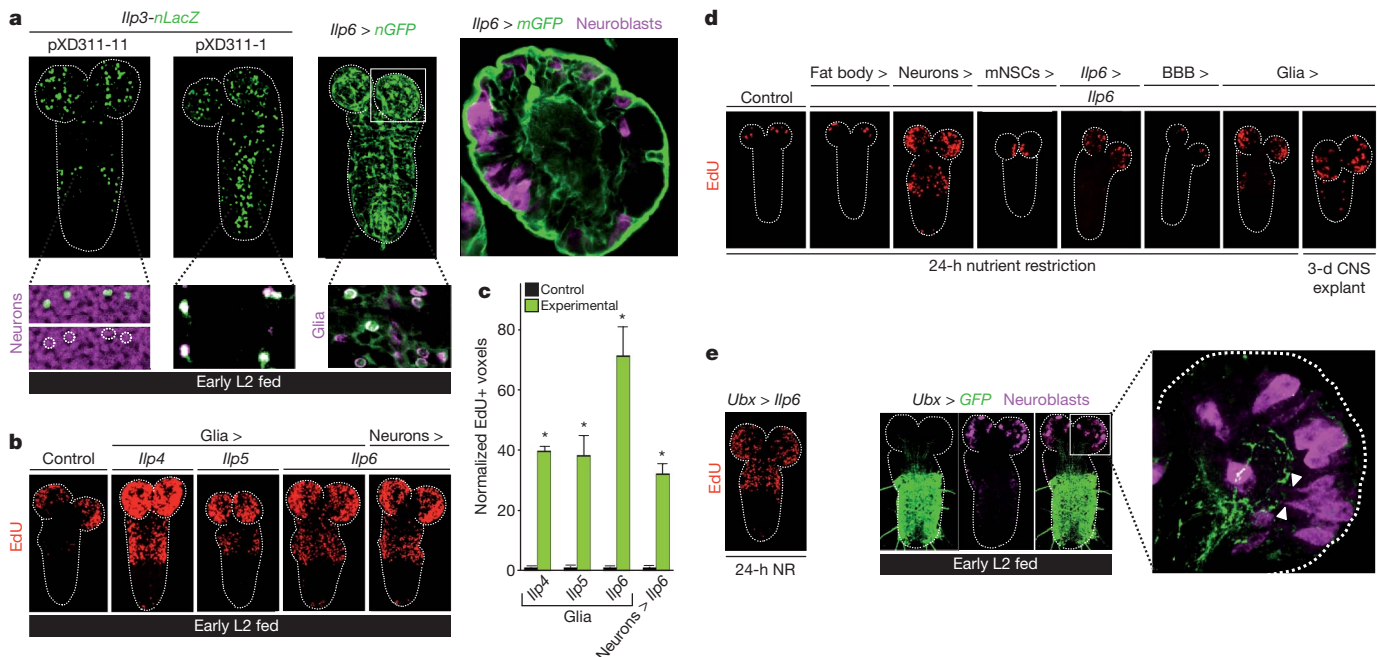
distinct from the *Ilp2*<sup>+</sup> mNSCs (Supplementary Fig. 4). Two different *Ilp3-lacZ* transgenes<sup>17</sup> indicate that *Ilp3* is expressed in some glia (Repo<sup>+</sup> cells) and neurons (Elav<sup>+</sup> cells). An *Ilp6-GAL4* insertion (see Methods) indicates that *Ilp6* is also expressed in glia, including the cortex glia surrounding neuroblasts and the glia of the blood-brain barrier (BBB) (Fig. 3a).

We next assessed the ability of each of the seven ILPs to reactivate neuroblasts when overexpressed in glia or in neurons (Supplementary Table 2). Pan-glial or pan-neuronal overexpression of ILP4, 5 or 6 led to precocious reactivation under fed conditions (Fig. 3b, c). Each of these manipulations also bypassed the nutrient restriction block to neuroblast reactivation, as did overexpression of ILP2 in glia or in neurons, or ILP3 in neurons (Fig. 3d and Supplementary Table 2). In all of these ILP overexpressions, and even when ILP6 was expressed in the posterior *Ultrabithorax* domain (Fig. 3e), the temporal rather than the spatial pattern of reactivation was affected. Importantly, experiments blocking cell signalling with SHI<sup>DN</sup> indicate that glia rather than neurons are critical for neuroblast reactivation (Fig. 4a, b). Interestingly, glial-specific overexpression of ILP3–6 did not significantly alter larval mass (Fig. 2c). Thus, in contrast to mNSC-derived ILPs, glial-derived ILPs promote CNS growth without affecting body growth.



**Figure 2 | Insulin-like peptides but not mNSCs control neuroblast reactivation.** **a**, EdU-labelled CNSs from various *Ilp* or *InR* mutants show decreased reactivation whereas larvae with *Ilp2-GAL4* driving *UAS-p60* (mNSCs > *p60*) do not. **b**, EdU incorporation time course in the CNS of

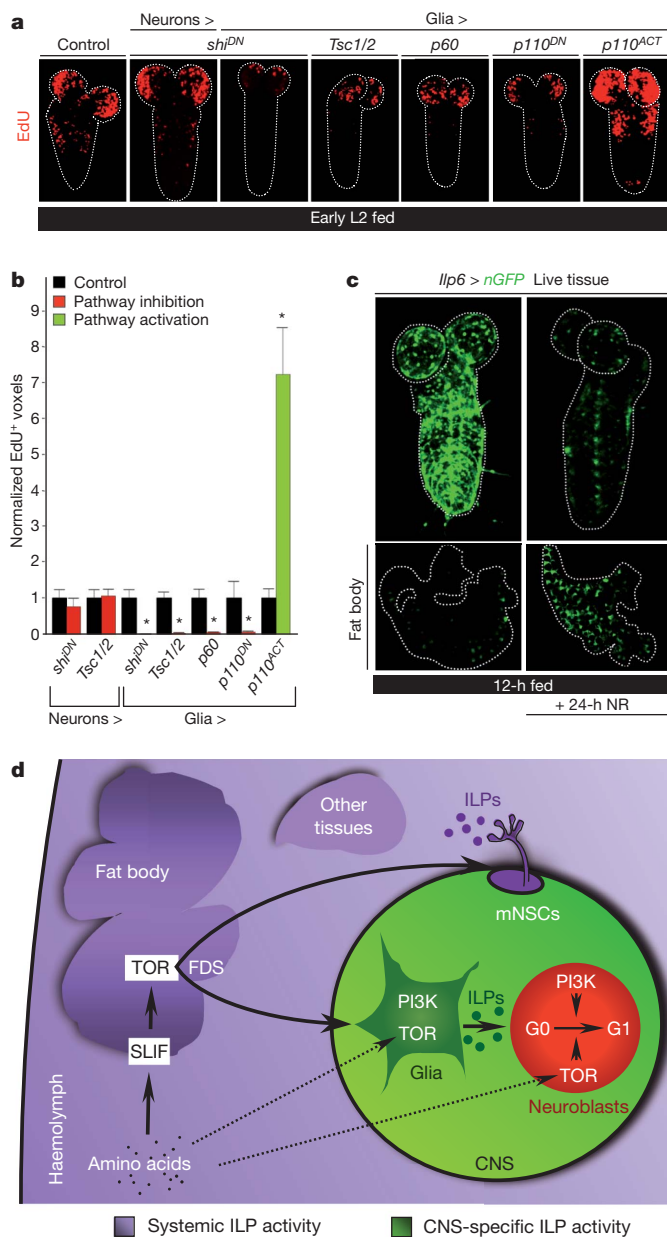
*Df(3L)Ilp1-5* larvae. **c**, The mass of fed L3 larvae at the wandering (W) stage is significantly altered by *Ilp2-GAL4* (mNSCs >) driving PI3K signalling components but not by *repo-GAL4* (Glia >) driving *Ilp* genes.



**Figure 3 | CNS-specific Insulin-like peptides are sufficient for neuroblast reactivation.** **a**, Panels show expression of *Ilp3-nLacZ* in subsets of neurons (XD311-11) and glia (XD311-1) and *Ilp6-GAL4* (*Ilp6 > nGFP* and *Ilp6 > mGFP*) in glia, including BBB and cortex glia. **b**, **d**, EdU-labelled CNSs from larvae overexpressing *Ilp* genes in various cell types (see Methods for *GAL4* drivers used). **c**, Histograms of normalized EdU<sup>+</sup> voxels in the thoracic CNS for the genotypes in **b**. **e**, *Ilp6* overexpression in the *Ultrabithorax* domain

(*Ubx > Ilp6*) reactivates neuroblasts in the normal spatial pattern during nutrient restriction (NR; left panel). Quiescent/enlarging neuroblasts in the central brain, far from the posterior *Ubx* domain (middle panels), extend cytoplasmic processes (arrowheads) towards the neuropil, close to long *Ubx*<sup>+</sup> cell processes (right panel). The range of *Ilp6* activity is difficult to determine from this experiment. Neurons, glia and neuroblasts are marked by *Elav*, *Repo* and *Miranda* respectively.





**Figure 4** | *Ilp6*-expressing glia are nutritionally regulated. **a**, EdU-labelled CNSs from larvae expressing the components indicated. **b**, Histograms of normalized EdU<sup>+</sup> voxels in the thoracic CNS for genotypes in **a**. **c**, *Ilp6 > nGFP* expression in the CNS and fat body of fed versus nutrient-restricted (NR) larvae. **d**, Relay model for amino-acid-dependent fat body regulation of CNS and body growth. CNS-restricted (green) and systemic (purple) pools of Insulin-like peptides (ILPs) are functionally segregated. Direct amino-acid sensing by glia and neuroblasts may contribute to neuroblast reactivation (dashed arrows). See text for details.

Focusing on ILP6, we used CNS explant cultures to demonstrate directly that glial overexpression was sufficient to substitute for the FDS during neuroblast exit from quiescence (Fig. 3d). *In vivo*, ILP6 was sufficient to induce reactivation during nutrient restriction when overexpressed via its own promoter or specifically in cortex glia but not in the subperineurial BBB glia, nor in many other CNS cells that we tested (Fig. 3d and Supplementary Table 1). Hence, cortex glia possess the appropriate processing machinery and/or location to deliver reactivating ILP6 to neuroblasts. *Ilp6* mRNA is known to be up- rather than downregulated in the larval fat body during starvation<sup>19</sup> and, accordingly, *Ilp6*-GAL4 activity is increased in this tissue after nutrient restriction (Fig. 4c). Conversely, we found that *Ilp6*-GAL4 is strongly

downregulated in CNS glia during nutrient restriction (Fig. 4c). Thus, dietary nutrients stimulate glia to express *Ilp6* at the transcriptional level. Consistent with this, an important transducer of nutrient signals, the TOR/PI3K network, is necessary and sufficient in glia (but not in neurons) for neuroblast reactivation (Fig. 4a, b). Together, the genetic and expression analyses indicate that nutritionally regulated glia relay the FDS to quiescent neuroblasts via ILPs.

This study used an integrative physiology approach to identify the relay mechanism regulating a nutritional checkpoint in neural progenitors. A central feature of the fat-body–glia–neuroblasts relay model is that glial insulin signalling bridges the amino-acid/TOR-dependent FDS with InR/PI3K/TOR signalling in neuroblasts (Fig. 4d). The importance of glial ILP signalling during neuroblast reactivation is also underscored by an independent study, published while this work was under revision<sup>27</sup>. As TOR signalling is also required in neuroblasts and glia, direct amino-acid sensing by these cell types may also impinge upon the linear tissue relay. This would then constitute a feed-forward persistence detector<sup>28</sup>, ensuring that neuroblasts exit quiescence only if high amino-acid levels are sustained rather than transient. We also showed that the CNS ‘compartment’ in which glial ILPs promote growth is functionally isolated, perhaps by the BBB, from the systemic compartment where mNSC ILPs regulate the growth of other tissues. The existence of two functionally separate ILP pools may explain why bovine insulin cannot reactivate neuroblasts in CNS organ culture<sup>14</sup>, despite being able to activate *Drosophila* InR *in vitro*<sup>29</sup>. Given that insulin/PI3K/TOR signalling components are highly conserved between insects and vertebrates, it will be important to address whether mammalian adipose or hepatic tissues signal to glia and whether or not this involves an insulin/IGF relay to CNS progenitors. In this regard, it is intriguing that brain-specific overexpression of IGF1 can stimulate cell-cycle re-entry of mammalian cortical neural progenitors<sup>30</sup>, indicating utilization of at least part of the mechanism identified here in *Drosophila*.

## METHODS SUMMARY

For GAL4/UAS experiments, *Drosophila* were raised at 29 °C unless otherwise stated. Larvae hatching within a 2 h window were transferred to cornmeal food (5.9% glucose, 6.6% cornmeal, 1.2% baker’s yeast, 0.7% agar in water) or nutrient-restricted medium (5% sucrose, 1% agar in PBS) and further synchronized by selecting L2 larvae morphologically from an L1/L2 moulting population. For EdU experiments, dissected CNSs were incubated for 1 h in 10 μM EdU/PBS, fixed for 15 min in 4% formaldehyde/PBS and Alexa Fluor azide was detected according to instructions (Click-iT EdU Imaging Kit, Invitrogen). CNS explants were cultured on 8-μm pore-size inserts in Schneider’s medium, 10% fetal calf serum, 2 mM L-glutamine (Gibco) and 1× Pen Strep (Gibco) in 24-well Transwell plates (Costar) in a humidified chamber at 25 °C. For EdU quantifications, the ‘thoracic’ region used corresponds to the ventral nerve cord from the level of the brain lobes down to A1/A2. EdU<sup>+</sup> voxels were quantified using Volocity (Improvision) from an average of ten CNSs per experimental genotype, normalized to controls processed in parallel (siblings or half-siblings), using Leica SP5 scans (LAS AF software) with a 1.5-μm-step z-series. For larval mass measurements, triplicates of ~50 wandering L3 male larvae per genotype were transferred to pre-weighed microfuge tubes and wet weights determined using a Precisa XB 120A balance. For all histograms, error bars represent the s.e.m. and *P* values are from two-tailed Student’s *t*-tests with equal sample variance. Further details can be found in Methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 16 June 2010; accepted 24 January 2011.

Published online 23 February 2011.

1. Dhawan, J. & Rando, T. A. Stem cells in postnatal myogenesis: molecular mechanisms of satellite cell quiescence, activation and replenishment. *Trends Cell Biol.* **15**, 666–673 (2005).
2. Collier, H. A. What’s taking so long? S-phase entry from quiescence versus proliferation. *Nature Rev. Mol. Cell Biol.* **8**, 667–670 (2007).
3. Yanagida, M. Cellular quiescence: are controlling genes conserved? *Trends Cell Biol.* **19**, 705–715 (2009).
4. Chen, E. & Finkel, T. The tortoise, the hare, and the FoxO. *Cell Stem Cell* **5**, 451–452 (2009).

5. Sánchez-García, I., Vicente-Duenas, C. & Cobaleda, C. The theoretical basis of cancer-stem-cell-based therapeutics of cancer: can it be put into practice? *Bioessays* **29**, 1269–1280 (2007).
6. Betschinger, J. & Knoblich, J. A. Dare to be different: asymmetric cell division in *Drosophila*, *C. elegans* and vertebrates. *Curr. Biol.* **14**, R674–R685 (2004).
7. Egger, B., Chell, J. M. & Brand, A. H. Insights into neural stem cell biology from flies. *Phil. Trans. R. Soc. Lond. B* **363**, 39–56 (2008).
8. Doe, C. Q. Neural stem cells: balancing self-renewal with differentiation. *Development* **135**, 1575–1587 (2008).
9. Sousa-Nunes, R., Cheng, L. Y. & Gould, A. P. Regulating neural proliferation in the *Drosophila* CNS. *Curr. Opin. Neurobiol.* **20**, 50–57 (2010).
10. Truman, J. W. & Bate, M. Spatial and temporal patterns of neurogenesis in the central nervous system of *Drosophila melanogaster*. *Dev. Biol.* **125**, 145–157 (1988).
11. Tsuji, T., Hasegawa, E. & Isshiki, T. Neuroblast entry into quiescence is regulated intrinsically by the combined action of spatial Hox proteins and temporal identity factors. *Development* **135**, 3859–3869 (2008).
12. Kambadur, R. *et al.* Regulation of POU genes by castor and hunchback establishes layered compartments in the *Drosophila* CNS. *Genes Dev.* **12**, 246–260 (1998).
13. Isshiki, T., Pearson, B., Holbrook, S. & Doe, C. Q. *Drosophila* neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511–521 (2001).
14. Britton, J. S. & Edgar, B. A. Environmental control of the cell cycle in *Drosophila*: nutrition activates mitotic and endoreplicative cells by distinct mechanisms. *Development* **125**, 2149–2158 (1998).
15. Colombani, J. *et al.* A nutrient sensor mechanism controls *Drosophila* growth. *Cell* **114**, 739–749 (2003).
16. Brogiolo, W. *et al.* An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–221 (2001).
17. Ikeya, T., Galic, M., Belawat, P., Nairz, K. & Hafen, E. Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. *Curr. Biol.* **12**, 1293–1300 (2002).
18. Rulifson, E. J., Kim, S. K. & Nüsse, R. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science* **296**, 1118–1120 (2002).
19. Slaidina, M., Delanoue, R., Gronke, S., Partridge, L. & Leopold, P. A *Drosophila* insulin-like peptide promotes growth during nonfeeding states. *Dev. Cell* **17**, 874–884 (2009).
20. Okamoto, N. *et al.* A fat body-derived IGF-like peptide regulates postfeeding growth in *Drosophila*. *Dev. Cell* **17**, 885–891 (2009).
21. Géminard, C., Rulifson, E. J. & Leopold, P. Remote control of insulin secretion by fat cells in *Drosophila*. *Cell Metab.* **10**, 199–207 (2009).
22. Polak, P. & Hall, M. N. mTOR and the control of whole body metabolism. *Curr. Opin. Cell Biol.* **21**, 209–218 (2009).
23. Neufeld, T. P. Body building: regulation of shape and size by PI3K/TOR signaling during development. *Mech. Dev.* **120**, 1283–1296 (2003).
24. Teleman, A. A. Molecular mechanisms of metabolic regulation by insulin in *Drosophila*. *Biochem. J.* **425**, 13–26 (2010).
25. Grönke, S., Clarke, D. F., Broughton, S., Andrews, T. D. & Partridge, L. Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet.* **6**, e1000857 (2010).
26. Zhang, H. *et al.* Deletion of *Drosophila* insulin-like peptides causes growth defects and metabolic abnormalities. *Proc. Natl Acad. Sci. USA* **106**, 19617–19622 (2009).
27. Chell, J. M. & Brand, A. H. Nutrition-responsive glia control exit of neural stem cells from quiescence. *Cell* **143**, 1161–1173 (2010).
28. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl Acad. Sci. USA* **100**, 11980–11985 (2003).
29. Fernandez-Almonacid, R. & Rosen, O. M. Structure and ligand specificity of the *Drosophila melanogaster* insulin receptor. *Mol. Cell. Biol.* **7**, 2718–2727 (1987).
30. Hodge, R. D., D'Ercole, A. J. & O'Kusky, J. R. Insulin-like growth factor-I accelerates the cell cycle by decreasing G1 phase length and increases cell cycle reentry in the embryonic cerebral cortex. *J. Neurosci.* **24**, 10201–10210 (2004).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to A. Brand, S. Cohen, B. Edgar, U. Gaul, E. Hafen, C. Klambt, T. Lee, S. Leever, P. Leopold, F. Matsuzaki, I. Miguel-Aliaga, T. Neufeld, R. Palmer, L. Partridge, L. Pick, E. Sanchez-Herrero, H. Stocker, N. Tapon and T. Xu, and also to the Bloomington stock centre and Kyoto National Institute of Genetics (NIG) for *Drosophila* stocks, antibodies and plasmids. We also acknowledge I. Salecker, J.-P. Vincent, A. Bailey, E. Cinnamon, L. Cheng, R. Makki, A. Matheu, P. Pachnis, P. Serpente and I. Stefana for providing advice, reagents and critical reading of the manuscript. The authors were supported by the Medical Research Council (U117584237).

**Author Contributions** R.S.-N. and A.P.G. designed the experiments, R.S.-N. and L.L.Y. performed the experiments and R.S.-N. and A.P.G. wrote the manuscript. All authors have read and subscribe to the contents of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.P.G. ([agould@nimr.mrc.ac.uk](mailto:agould@nimr.mrc.ac.uk)).

## METHODS

**Rearing and staging of *Drosophila* larvae.** To assist larval genotyping, lethal chromosomes were re-established over *Dfd-YFP* balancers. For EdU experiments, crosses were performed in cages with grape-juice plates (25% (v/v) grape-juice, 1.25% (w/v) sucrose, 2.5% (w/v) agar) supplemented with live yeast paste. For GAL4/UAS experiments, larvae hatched within a 2-h time window were transferred to our standard cornmeal food (5.9% w/v glucose, 6.6% cornmeal, 1.2% baker's yeast, 0.7% agar in water) or to nutrient-restricted (NR) medium (5% sucrose, 1% agar in PBS). *Drosophila* were raised at 29 °C throughout, with the following exceptions owing to lethality at high temperature: *tub-GAL80<sup>ts</sup>*, *repo > shi<sup>DN</sup>* and *tub-GAL80<sup>ts</sup>, repo > Ilp2* *Drosophila* were raised at 25 °C during embryogenesis and 29 °C during larval development, other *Ilp2* overexpressions were performed at 25 °C throughout. At the time of dissection, development was further synchronized by selecting L2 animals morphologically from a mixed L1/L2 moulting population. *Df(3L)Ilp2-3*, *Ilp5<sup>3</sup>* and *Df(3L)Ilp1-5* homozygotes develop considerably slower than controls, so EdU-incorporation experiments used morphological staging after the L1/L2 and L2/L3 moults. Co-expression of *Dcr-2* was used to enhance knockdown efficiency for the *slif<sup>ANTI</sup>* allele, in which antisense *slif* sequences are under UAS control from an EP element insertion. As absolute numbers of reactivated neuroblasts can vary with small differences in temperature and humidity, parallel control experiments were carried out for each genetic background (using the siblings or half-siblings of experimental animals), rather than using a single control.

***Drosophila* strains.** Stocks used in this study were: *Tor<sup>2L19</sup>* (ref. 31), *InR<sup>31</sup>* (ref. 32), *Df(3L)Ilp1-5*, *Df(X)Ilp6* and *Df(X)Ilp7* (ref. 33), *Ilp1<sup>1</sup>*, *Ilp2<sup>1</sup>*, *Ilp3<sup>1</sup>*, *Ilp4<sup>1</sup>*, *Ilp5<sup>1</sup>*, *Ilp5<sup>3</sup>*, *Df(X)Ilp6<sup>41</sup>*, *Df(X)Ilp6<sup>68</sup>*, *Ilp7<sup>1</sup>*, *Df(3L)Ilp2-3*, *Df(3L)Ilp1-4* (ref. 34), *slif<sup>ANTI</sup>* (ref. 35), *UAS-Tsc1/2* (ref. 36), *UAS-Rheb*, *UAS-InR<sup>DN</sup>* = *UAS-InR<sup>K1409A</sup>*, *UAS-InR<sup>ACT</sup>* = *UAS-InR<sup>A1325D</sup>*, FB driver = *Cg-GAL4* (ref. 37), pan-glia driver = *repo-GAL4* (ref. 38), *OK107-GAL4* (ref. 39), *eg-GAL4* (ref. 40), *DopR-GAL4* (ref. 41), *btl-GAL4* (ref. 42), *UAS-CD8::GFP*, *FRT82B*, *tub-GAL80<sup>ts</sup>*, *Sco/CyO*, *Dfd-YFP* and *Dr/TM6B*, *Sb*, *Dfd-YFP* (Bloomington *Drosophila* Stock Center), *UAS-p110<sup>DN</sup>* = *UAS-p110<sup>A2860C</sup>* and *UAS-p110<sup>ACT</sup>* = *UAS-p110<sup>CAAX</sup>* (ref. 43), *UAS-Pten* (ref. 44), *UAS-p60* (ref. 45), *UAS-Pdk<sup>ACT</sup>* = *UAS-Pdk1<sup>A467V</sup>* (ref. 46), *UAS-foxo* (ref. 47), *UAS-Dcr2* (VDR), *UAS-Ilp7* (ref. 48), *UAS-Ilp1*, *UAS-Ilp2*, *UAS-Ilp3*, *UAS-Ilp4*, *UAS-Ilp5*, *UAS-Ilp6*, *Ilp3-nLacZ<sup>2pX311-1</sup>* and *Ilp3-nLacZ<sup>2pX311-11</sup>* (both recapitulate endogenous *Ilp3* expression in L3 mNSCs, ref. 49), *UAS-shi<sup>DN</sup>* (ref. 50), NB driver = *nab-GAL4<sup>NP4604</sup>* (ref. 51), cortex glia driver = *NP577-GAL4* and ensheathing glia driver = *NP6520-GAL4* (ref. 52), *Ilp6-GAL4* = *NP1079-GAL4* (NIG stock centre), subperineurial BBB glia driver = *moody-GAL4* (ref. 53), midline glia driver = *slit-GAL4* (ref. 54), midline glia/neuronal driver = *sim-GAL4* (ref. 55), mNSC driver = *Ilp2-GAL4* (ref. 56), pan-neuronal driver = *n-syb-GAL4* (ref. 57), *Ubx-GAL4* (ref. 58), *wg-GAL4* (ref. 59), *en-GAL4* was a gift from A. Brand via J.-P. Vincent, *repo-FLP* (ref. 60).

**EdU detection, immunostaining, *in situ* hybridization and imaging.** L1 and L2 tissues were immobilized on poly-L-Lysine-coated slides for all stainings, except for CNS explants. For EdU experiments, dissected CNSs were incubated for 1 h in 10 μM EdU/PBS, fixed for 15 min in 4% formaldehyde/PBS, followed by detection of Alexa Fluor azide according to the manufacturer's instructions (Click-iT EdU Imaging Kit, Invitrogen) and washing in 0.1% Triton/PBS. Antibody staining and *in situ* hybridization were performed according to standard protocols. Primary antibodies used in this study were: rabbit anti-β-Galactosidase (Molecular Probes) 1/2,000; rabbit anti-GFP (Invitrogen) 1/1,000; mouse anti-Repo 1/20; mouse anti-Miranda 1/20 and rat anti-Elav 1/100 (Developmental Studies Hybridoma Bank); pre-adsorbed alkaline-phosphatase-conjugated sheep anti-digoxigenin (Roche) 1/2,000. Secondary antibodies used were: F(ab')<sub>2</sub> fragments conjugated to either Alexa-Fluor-488, Alexa-Fluor-633 (Molecular Probes) or Cy3 (Jackson), used at 1/250–1/2,000. Live tissues were photographed in PBS. Fixed tissues labelled for fluorescence microscopy were mounted in Vectashield (Vector Laboratories) whereas those processed for *in situ* hybridization were mounted in 80% glycerol. Fluorescent images were acquired with a Leica SP5 confocal microscope (LAS AF software) and bright-field images were acquired with a Zeiss Axiophot2 microscope (AxioVision software). Images of the whole CNS are projections of a 1.5-μm-step z-series. Images of fat body and of high-magnification double-labels of parts of the CNS are single sections except for the right panel of Fig. 3e, which is a projection of 13 sections from a z-series.

**CNS explant cultures.** Explanted CNSs from larvae hatched within a 2-h window were cultured for 3–4 days on 8-μm pore-size inserts in 10 μM EdU in Schneider's medium, 10% fetal calf serum, 2 mM L-glutamine (Gibco) and 1× Pen Strep (Gibco), in 24-well Transwell plates (Costar) placed in a humidified chamber at 25 °C.

**Quantification of EdU incorporation.** The 'thoracic' region used for EdU quantifications corresponds to the ventral nerve cord from brain-lobe level down to A1/A2, distinguishable from more posterior neuromeres by a sharp transition in neuroblast density (Fig. 1a). The numbers of EdU<sup>+</sup> voxels per CNS were determined using Volocity (Improvision) from Leica SP5 confocal microscope scans

(LAS AF software) using a 1.5-μm-step z-series. An average of ten CNSs were quantified per experimental genotype and controls (siblings or half-siblings) were processed in parallel. Control and experimental values were normalized using the average number of control EdU<sup>+</sup> voxels. For all histograms, error bars represent standard error of the mean (s.e.m.) of normalized values and asterisks indicate *P* < 0.05 using two-tailed Student's *t*-tests with equal sample variance.

**Larval mass measurements.** Wet weights were determined for wandering L3 male larvae, sexed and genotyped in PBS, dabbed dry with tissue and transferred to pre-weighed microfuge tubes. For each data point, triplicate samples, each containing an average of 50 animals per genotype were weighed (Precisa XB 120A balance).

- Oldham, S. *et al.* Genetic and biochemical characterization of dTOR, the *Drosophila* homolog of the target of rapamycin. *Genes Dev.* **14**, 2689–2694 (2000).
- Brogiolo, W. *et al.* An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213–221 (2001).
- Zhang, H. *et al.* Deletion of *Drosophila* insulin-like peptides causes growth defects and metabolic abnormalities. *Proc. Natl Acad. Sci. USA* **106**, 19617–19622 (2009).
- Gronke, S. *et al.* Molecular evolution and functional characterization of *Drosophila* insulin-like peptides. *PLoS Genet.* **6**, e1000857 (2010).
- Colombani, J. *et al.* A nutrient sensor mechanism controls *Drosophila* growth. *Cell* **114**, 739–749 (2003).
- Tapon, N. *et al.* The *Drosophila* tuberous sclerosis complex gene homologs restrict cell growth and cell proliferation. *Cell* **105**, 345–355 (2001).
- Hennig, K. M., Colombani, J. & Neufeld, T. P. TOR coordinates bulk and targeted endocytosis in the *Drosophila melanogaster* fat body to regulate cell growth. *J. Cell Biol.* **173**, 963–974 (2006).
- Xiong, W. C. *et al.* repo encodes a glial-specific homeo domain protein required in the *Drosophila* nervous system. *Genes Dev.* **8**, 981–994 (1994).
- Connolly, J. B. *et al.* Associative learning disrupted by impaired G<sub>s</sub> signaling in *Drosophila* mushroom bodies. *Science* **274**, 2104–2107 (1996).
- Ito, K., Urban, J. & Technau, G. M. Distribution, classification, and development of *Drosophila* glial cells in the late embryonic and early larval ventral nerve cord. *Roux Arch. Dev. Biol.* **204**, 284–307 (1995).
- Hacker, U. *et al.* piggyBac-based insertional mutagenesis in the presence of stably integrated P elements in *Drosophila*. *Proc. Natl Acad. Sci. USA* **100**, 7720–7725 (2003).
- Shiga, Y., Tanaka-Matajats, M. & Hayashi, S. A nuclear GFP/β-galactosidase fusion protein as a marker for morphogenesis in living *Drosophila*. *Dev. Growth Differ.* **38**, 99–106 (1996).
- Leesters, S. J. *et al.* The *Drosophila* phosphoinositide 3-kinase Dp110 promotes cell growth. *EMBO J.* **15**, 6584–6594 (1996).
- Huang, H. *et al.* PTEN affects cell size, cell proliferation and apoptosis during *Drosophila* eye development. *Development* **126**, 5365–5372 (1999).
- Weinkove, D. *et al.* Regulation of imaginal disc cell size, cell number and organ size by *Drosophila* class Ia phosphoinositide 3-kinase and its adaptor. *Curr. Biol.* **9**, 1019–1029 (1999).
- Rintelen, F., Stocker, H., Thomas, G. & Hafen, E. PDK1 regulates growth through Akt and S6K in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15020–15025 (2001).
- Puig, O., Marr, M. T., Ruhf, M. L. & Tjian, R. Control of cell number by *Drosophila* FOXO: downstream and feedback regulation of the insulin receptor pathway. *Genes Dev.* **17**, 2006–2020 (2003).
- Miguel-Alíaga, I., Thor, S. & Gould, A. P. Postmitotic specification of *Drosophila* insulinergic neurons from pioneer neurons. *PLoS Biol.* **6**, e58 (2008).
- Ikeya, T. *et al.* Nutrient-dependent expression of insulin-like peptides from neuroendocrine cells in the CNS contributes to growth regulation in *Drosophila*. *Curr. Biol.* **12**, 1293–1300 (2002).
- Moline, M. M., Southern, C. & Bejsovec, A. Directionality of wingless protein transport influences epidermal patterning in the *Drosophila* embryo. *Development* **126**, 4375–4384 (1999).
- Maurange, C., Cheng, L. & Gould, A. P. Temporal transcription factors and their targets schedule the end of neural proliferation in *Drosophila*. *Cell* **133**, 891–902 (2008).
- Awasaki, T., Lai, S.-L., Ito, K. & Lee, T. Organization and postembryonic development of glial cells in the adult central brain of *Drosophila*. *J. Neurosci.* **28**, 13742–13753 (2008).
- Schwabe, T. *et al.* GPCR Signaling is required for blood-brain barrier formation in *Drosophila*. *Cell* **123**, 133–144 (2005).
- Scholz, H., Sadlowski, E., Klaes, A. & Klambt, C. Control of midline glia development in the embryonic *Drosophila* CNS. *Mech. Dev.* **64**, 139–151 (1997).
- Kidd, T., Bland, K. S. & Goodman, C. S. Slit is the midline repellent for the Robo receptor in *Drosophila*. *Cell* **96**, 785–794 (1999).
- Rulifson, E. J., Kim, S. K. & Nusse, R. Ablation of insulin-producing neurons in flies: growth and diabetic phenotypes. *Science* **296**, 1118–1120 (2002).
- Pospisilik, J. A. *et al.* *Drosophila* genome-wide obesity screen reveals hedgehog as a determinant of brown versus white adipose cell fate. *Cell* **140**, 148–160 (2010).
- de Navas, L. F., Garaulet, D. L. & Sanchez-Herrero, E. The *Ultrathorax* Hox gene of *Drosophila* controls haltere size by regulating the Dpp pathway. *Development* **133**, 4495–4506 (2006).
- Pfeiffer, S., Alexandre, C., Calleja, M. & Vincent, J. P. The progeny of wingless-expressing cells deliver the signal at a distance in *Drosophila* embryos. *Curr. Biol.* **10**, 321–324 (2000).
- Silies, M. *et al.* Glial cell migration in the eye disc. *J. Neurosci.* **27**, 13130–13139 (2007).

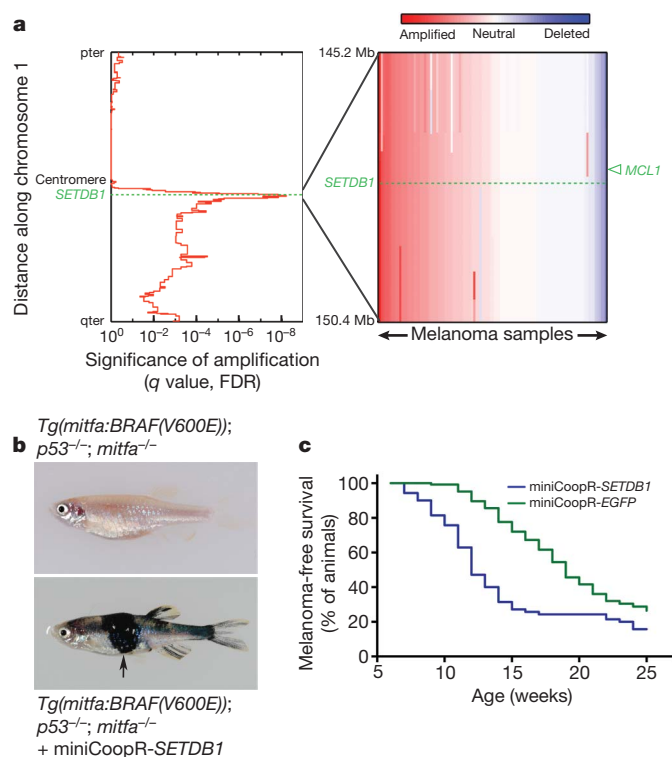


# The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset

Craig J. Ceol<sup>1†\*</sup>, Yariv Houvras<sup>1,2†\*</sup>, Judit Jane-Valbuena<sup>3,4</sup>, Steve Bilodeau<sup>5</sup>, David A. Orlando<sup>5</sup>, Valentine Battisti<sup>6</sup>, Lauriane Fritsch<sup>6</sup>, William M. Lin<sup>3,4</sup>, Travis J. Hollmann<sup>7</sup>, Fabrizio Ferré<sup>8</sup>, Caitlin Bourque<sup>1</sup>, Christopher J. Burke<sup>1</sup>, Laura Turner<sup>1</sup>, Audrey Uong<sup>1</sup>, Laura A. Johnson<sup>3,4</sup>, Rameen Beroukhi<sup>3,4</sup>, Craig H. Mermel<sup>3,4</sup>, Massimo Loda<sup>7</sup>, Slimane Ait-Si-Ali<sup>6</sup>, Levi A. Garraway<sup>3,4</sup>, Richard A. Young<sup>5</sup> & Leonard I. Zon<sup>1</sup>

The most common mutation in human melanoma, *BRAF(V600E)*, activates the serine/threonine kinase BRAF and causes excessive activity in the mitogen-activated protein kinase pathway<sup>1,2</sup>. *BRAF(V600E)* mutations are also present in benign melanocytic naevi<sup>3</sup>, highlighting the importance of additional genetic alterations in the genesis of malignant tumours. Such changes include recurrent copy number variations that result in the amplification of oncogenes<sup>4,5</sup>. For certain amplifications, the large number of genes in the interval has precluded an understanding of the cooperating oncogenic events. Here we have used a zebrafish melanoma model to test genes in a recurrently amplified region of chromosome 1 for the ability to cooperate with *BRAF(V600E)* and accelerate melanoma. SETDB1, an enzyme that methylates histone H3 on lysine 9 (H3K9), was found to accelerate melanoma formation significantly in zebrafish. Chromatin immunoprecipitation coupled with massively parallel DNA sequencing and gene expression analyses uncovered genes, including *HOX* genes, that are transcriptionally dysregulated in response to increased levels of SETDB1. Our studies establish *SETDB1* as an oncogene in melanoma and underscore the role of chromatin factors in regulating tumorigenesis.

To identify genes that promote melanoma, we focused on genomic regions that are subject to copy number amplification in human tumour samples. In a study of 101 cell lines and short-term cultures of melanoma cells, chromosome 1q21 (chr1: 147.2–149.2 megabases) was identified as a recurrently amplified interval<sup>6</sup> (Fig. 1a). The same region was implicated in another comprehensive analysis of copy number variation in melanoma<sup>4</sup>. To test candidate genes from this interval for the ability to accelerate melanoma, we developed an assay in transgenic (Tg) zebrafish in which *BRAF(V600E)* is expressed under the control of a melanocyte-specific gene (*mitfa*) promoter on a *p53* (also known as *tp53*) mutant background (*p53*<sup>-/-</sup>) (Supplementary Fig. 1). Melanomas and melanocytes that develop in *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> zebrafish<sup>7</sup> are suppressed by a *mitfa*<sup>-/-</sup> mutation. We engineered a transposon-based vector called miniCoopR that rescues melanocytes and melanomas in a *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup>; *mitfa*<sup>-/-</sup> strain and drives the expression of a candidate gene in these rescued tissues. We identified genes that were present in the human 1q21 region and were overexpressed as messenger RNAs in 1q21-amplified melanomas based on Affymetrix microarrays. Candidate human genes were cloned into the miniCoopR vector and injected into one-cell stage *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup>; *mitfa*<sup>-/-</sup> zebrafish embryos. Tumour



**Figure 1 | SETDB1 accelerates melanoma formation in zebrafish.** **a**, Left, significance of copy number amplification along chromosome 1 in human melanoma samples, as assessed by using the algorithm GISTIC (genomic identification of significant targets in cancer)<sup>8,20</sup>. Significance values were determined by the false discovery rate (FDR) test. Right, copy number profiles in the human 1q21 interval in melanoma samples (vertical bars). The positions of *SETDB1* (dashed line) and *MCL1* (arrowhead) are indicated. Mb, megabase. **b**, The *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup>; *mitfa*<sup>-/-</sup> strain (top) was injected with miniCoopR-cloned candidate oncogenes. In animals injected with miniCoopR-*SETDB1* (bottom), the melanocytes are rescued, and melanomas (arrow) rapidly develop. **c**, Melanoma-free survival curves for *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup>; *mitfa*<sup>-/-</sup> zebrafish injected with miniCoopR-*SETDB1* (weighted average of 2 independent experiments, *n* = 70) or miniCoopR-*EGFP* (weighted average of 3 independent experiments, *n* = 125).

<sup>1</sup>Stem Cell Program and Hematology/Oncology, Children's Hospital Boston, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>2</sup>Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Departments of Medical Oncology, Cancer Biology, and Center for Cancer Genome

Discovery, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>4</sup>The Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. <sup>5</sup>Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. <sup>6</sup>UMR7216 Epigénétique et Destin Cellulaire, CNRS, Université Paris-Diderot, 35 rue Hélène Brion, 75013 Paris, France. <sup>7</sup>Center for Molecular Oncologic Pathology, Brigham and Women's Hospital, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>8</sup>A. Rossi Fanelli Biochemical Sciences Department, Sapienza University of Rome, Rome 00185, Italy. <sup>†</sup>Present addresses: Program in Molecular Medicine, Program in Cell Dynamics, and Department of Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA (C.J.C.); Departments of Surgery and Medicine, Weill Cornell Medical College and New York Presbyterian Hospital, New York, New York 10065, USA (Y.H.).

\*These authors contributed equally to this work.

incidence curves for the resultant adults showed that one gene in this interval, *SETDB1*, significantly accelerated melanoma onset ( $P = 9.4 \times 10^{-7}$ , logrank chi-squared test; Fig. 1b, c and Supplementary Fig. 2).

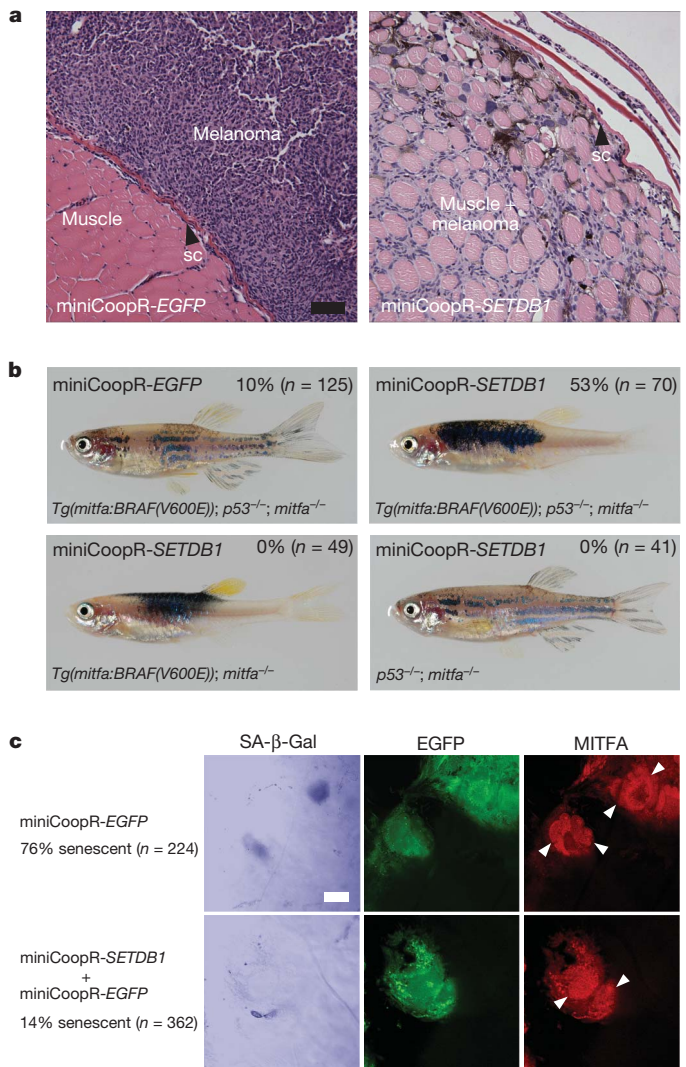
As in melanoma, *SETDB1* is focally amplified in non-small-cell lung cancer, small-cell lung cancer, ovarian cancer, hepatocellular carcinoma and breast cancer (Supplementary Fig. 3). The anti-apoptotic gene *MCL1* resides near *SETDB1* in the 1q21 interval, and knockdown of *MCL1* has been shown to diminish non-small-cell lung cancer proliferation and xenograft outgrowth<sup>8</sup>. However, *MCL1* is not overexpressed in the 1q21-amplified melanoma samples, so it was not tested in this study. No other gene accelerated the onset of melanomas, suggesting that *SETDB1* is a crucial gene that is amplified in the chromosome 1q21 interval. Using fluorescence *in situ* hybridization, we observed *SETDB1* amplification in short-term cultures of human melanoma cells (Supplementary Fig. 4), directly confirming the array-based copy number data from which our study originated.

Melanomas overexpressing *SETDB1* were more aggressive than tumours overexpressing enhanced green fluorescent protein (EGFP) when analysed at an equivalent stage and in the same *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>; mitfa<sup>-/-</sup>* genetic background. The melanomas expressing *SETDB1* were more locally invasive than the EGFP control melanomas (Fig. 2a; 94% (overexpressing *SETDB1*;  $n = 18$ ) versus 53% (expressing EGFP;  $n = 17$ ) of melanomas invaded the muscle ( $P = 1.6 \times 10^{-3}$ , Fisher's exact test), and 89% (*SETDB1*) versus 35% (EGFP) invaded the spinal column ( $P = 7.2 \times 10^{-3}$ , Fisher's exact test)). MiniCoopR-*SETDB1* transgenic melanomas had more extensive nuclear pleomorphism and larger nuclei than control tumours (Supplementary Fig. 5). By contrast, miniCoopR-*SETDB1* tumours showed similar levels of BRAF protein to control tumours, indicating that *SETDB1* did not accelerate melanoma formation by altering expression of the *BRAF(V600E)* transgene (Supplementary Fig. 6).

Melanocytes overexpressing *SETDB1* grew in confluent patches in zebrafish, unlike melanocytes in the EGFP-overexpressing control zebrafish, which grew in a wild-type stripe pattern. We analysed the genetic interactions that are responsible for these pigmentation differences. *SETDB1*-expressing melanocytes in the *Tg(mitfa:BRAF(V600E)); mitfa<sup>-/-</sup>* strain formed confluent patches, but *SETDB1*-expressing melanocytes in the *p53<sup>-/-</sup>; mitfa<sup>-/-</sup>* strain grew in a striped pattern (Fig. 2b). Although *SETDB1* and *BRAF(V600E)* cooperated to override normal pigment patterning, no tumours arose in miniCoopR-*SETDB1*-injected *Tg(mitfa:BRAF(V600E)); mitfa<sup>-/-</sup>* zebrafish, indicating that *SETDB1* overexpression does not have the same effect as loss of *p53* in tumour formation.

*BRAF(V600E)* induces senescence in human naevi and in cultured mammalian melanocytes<sup>9</sup>, and we suspected that the pigmentation differences might result from a failure of senescence and excess melanocyte proliferation caused by *SETDB1*. Using senescence-associated  $\beta$ -Galactosidase (SA- $\beta$ -Gal) staining<sup>10,11</sup>, we confirmed that *BRAF(V600E)* induces senescence in zebrafish melanocytes *in vivo* (Supplementary Fig. 7a–c). We stained miniCoopR-rescued melanocytes and found *SETDB1*-expressing melanocytes to be less senescent than those expressing only EGFP (Fig. 2c). *SETDB1*-expressing melanocytes also lacked the flattened morphology of senescent cells (Supplementary Fig. 7d). These results suggest that *SETDB1* overexpression may contribute to melanoma formation by abrogating oncogene-induced senescence.

To understand the gene expression changes that occur when *SETDB1* is overexpressed, we performed microarray analyses of zebrafish melanomas. We defined a gene signature comprising 67 human orthologues of genes that are downregulated in *SETDB1*-overexpressing zebrafish melanomas (Fig. 3a) and tested the relationship between this signature and *SETDB1* expression in human melanomas. Using gene set enrichment analysis (GSEA)<sup>12,13</sup>, we found that the gene signature was inversely correlated with *SETDB1* expression across a panel of 93 cell lines and short-term cultures of melanoma cells (Fig. 3b). *SETDB1*



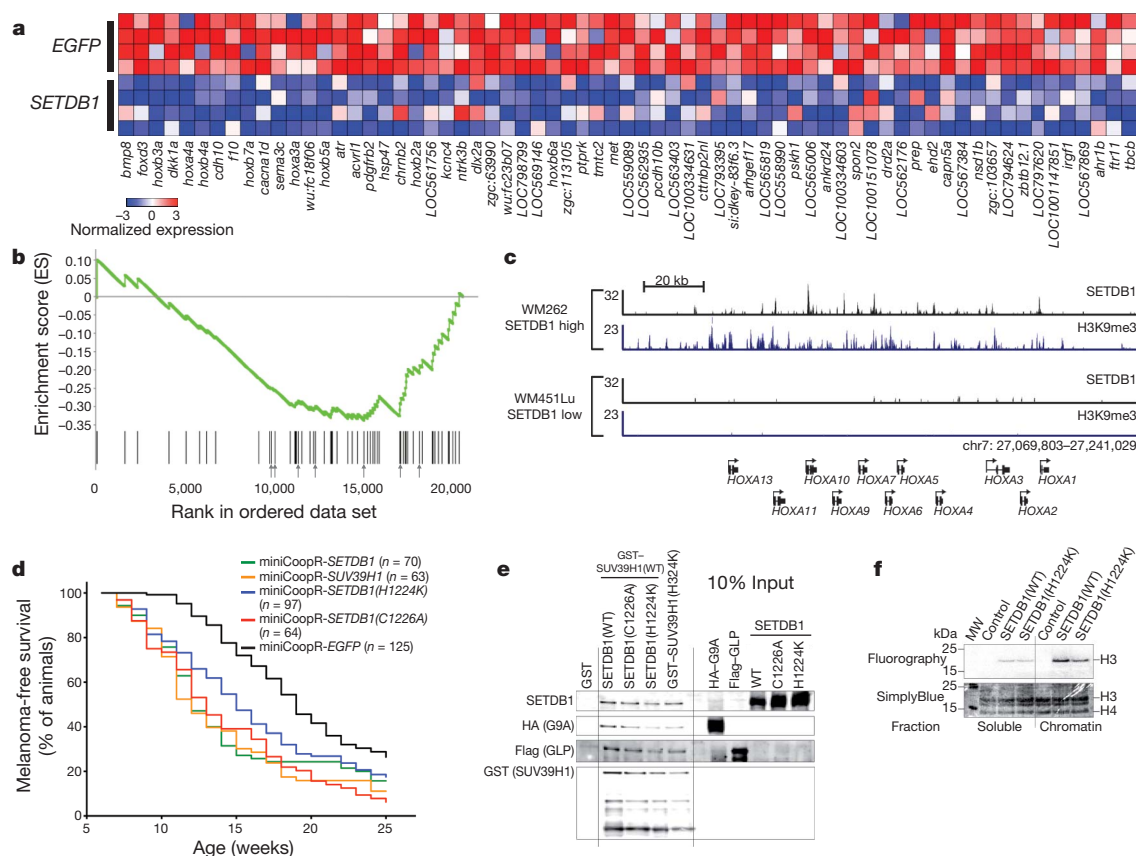
**Figure 2 | Effects of *SETDB1* on melanoma cells and melanocytes.**

**a**, Transverse sections of zebrafish melanomas at 2 weeks post onset of melanoma, visualized by staining with haematoxylin and eosin. At this time point, dorsal miniCoopR-EGFP melanomas (left) have exophytic growth, whereas miniCoopR-*SETDB1* melanomas (right) have invaded from the skin, through the collagen-rich stratum compactum (sc) of the dermis, into the underlying musculature. Scale bar, 70  $\mu$ m. **b**, *SETDB1* interacts with *BRAF(V600E)*, affecting the pigmentation pattern, but the *p53<sup>-/-</sup>* mutation is required for melanoma formation. miniCoopR-EGFP or miniCoopR-*SETDB1* was injected into the indicated transgenic strains. The photographs indicate pigmentation pattern differences before the time point at which melanomas begin to form in the *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>; mitfa<sup>-/-</sup>* background. Percentages indicate the melanoma incidence at 12 weeks of age;  $n$  = number of fish. **c**, *SETDB1* abrogates *BRAF(V600E)*-induced senescence. Left, brightfield pseudocoloured photomicrographs of SA- $\beta$ -Gal staining performed on scale-associated melanocytes. Centre and right, fluorescent photomicrographs of the same melanocytes. In this experiment, miniCoopR-rescued melanocytes express *mitfa*-promoter-driven EGFP (centre) and MITFA (right). Multiple nuclei (arrowheads) are present in *BRAF(V600E)*-expressing melanocytes. The percentage of senescent melanocytes is indicated ( $P = 7.3 \times 10^{-51}$ , chi-squared test);  $n$  = number of cells. Scale bar, 10  $\mu$ m.

overexpression led to a broad pattern of transcriptional changes, including conserved downregulation of a group of genes that is enriched for *HOX* genes and for transcriptional regulators.

To identify the direct targets of *SETDB1* across the genome in melanoma cells, we performed chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq). We identified *SETDB1* targets from WM262, a short-term culture of melanoma cells with high levels of *SETDB1* expression, and from WM451Lu, a short-term





**Figure 3 | SETDB1 target gene regulation and histone methyltransferase complex formation.** **a**, Heat-map of genes that are downregulated in zebrafish melanomas overexpressing SETDB1 compared with control (EGFP-expressing) melanomas. **b**, Graphical representation of the rank-ordered gene list derived from a panel of short-term cultures of human melanoma cells and stratified on the basis of SETDB1 expression level. The enrichment score is calculated based on a running-sum statistic, which increases when a gene (vertical line) in the gene set is encountered and decreases when one is not. GSEA shows that human orthologues of SETDB1-downregulated zebrafish genes are similarly downregulated in human melanomas as the level of SETDB1 increases (enrichment score (ES) = -0.37, normalized enrichment score (NES) = -1.47,  $q$  value = 0.034 (FDR test),  $P = 0.034$ ). The arrows indicate the positions of HOX genes. **c**, SETDB1 and H3K9me3 ChIP-seq profiles at the HOXA locus in human melanoma cells. The number of sequence reads is shown on the y axis. SETDB1 and overlapping H3K9me3 are present at the HOXA locus in WM262 cells but largely absent in WM451Lu cells, kb,

culture of melanoma cells with low levels of SETDB1 (Supplementary Fig. 8). These short-term cultures harbour the *BRAF(V600E)* mutation (Supplementary Fig. 9), and their proliferation is sensitive to changes in SETDB1 levels (Supplementary Figs 10 and 11). In murine embryonic stem cells, SETDB1 binds to the promoters of genes encoding developmental regulators, including *Hox* genes<sup>14</sup>. We observed differential binding of SETDB1 to genes in the *HOXA* cluster in melanoma cell lines with high and low levels of SETDB1 expression; SETDB1 is bound to *HOXA* genes in WM262 cells, whereas there is minimal binding in WM451Lu cells (Fig. 3c and Supplementary Tables 1 and 2). SETDB1 catalyses the trimethylation of histone H3 lysine 9 (H3K9me3), thereby promoting repression of its target genes. ChIP-seq for the H3K9me3 mark showed that H3K9me3 is present at the *HOXA* locus in WM262 cells but absent in WM451Lu cells (Fig. 3c). *HOX* gene expression is inversely correlated with SETDB1 levels in short-term cultures of melanoma cells (Fig. 3b), suggesting that enhanced target gene binding and repression may have a role in the SETDB1-mediated acceleration of melanoma onset. Additional ChIP-seq, for MCAF1 (also known as AM and ATF7IP; a methyltransferase-stimulatory cofactor of SETDB1)<sup>15</sup> in WM262 cells, suggests that the

effects of SETDB1 overexpression are mediated in part by MCAF1 (Supplementary Fig. 12). We assayed the effects of SETDB1 overexpression on target genes by infecting WM451Lu cells with a SETDB1-expressing lentivirus. Using SETDB1 ChIP-seq data from WM451Lu cells, we found that SETDB1-bound targets are significantly enriched in downregulated genes but not upregulated ones (Supplementary Fig. 13 and Supplementary Table 3), suggesting that a major consequence of SETDB1 amplification is repression of SETDB1-bound target genes. However, many SETDB1 target genes in both WM451Lu and WM262 short-term cultures of melanoma cells are not methylated, and additional analyses show a relationship between increasing SETDB1 levels and increasing expression of many SETDB1 target genes (Supplementary Fig. 14). To obtain a mechanistic insight into the role of SETDB1 in regulating gene expression, we undertook genetic and biochemical studies that evaluate methyltransferase activity. Recently, a complex containing SETDB1 and the H3K9 methyltransferases SUV39H1, G9A (also known as EHMT2) and GLP (also known as EHMT1) was discovered<sup>16</sup>. To examine the possibility that other methyltransferases act together with SETDB1 to modulate melanoma onset, we tested whether

effects of SETDB1 overexpression are mediated in part by MCAF1 (Supplementary Fig. 12).

We assayed the effects of SETDB1 overexpression on target genes by infecting WM451Lu cells with a SETDB1-expressing lentivirus. Using SETDB1 ChIP-seq data from WM451Lu cells, we found that SETDB1-bound targets are significantly enriched in downregulated genes but not upregulated ones (Supplementary Fig. 13 and Supplementary Table 3), suggesting that a major consequence of SETDB1 amplification is repression of SETDB1-bound target genes. However, many SETDB1 target genes in both WM451Lu and WM262 short-term cultures of melanoma cells are not methylated, and additional analyses show a relationship between increasing SETDB1 levels and increasing expression of many SETDB1 target genes (Supplementary Fig. 14).

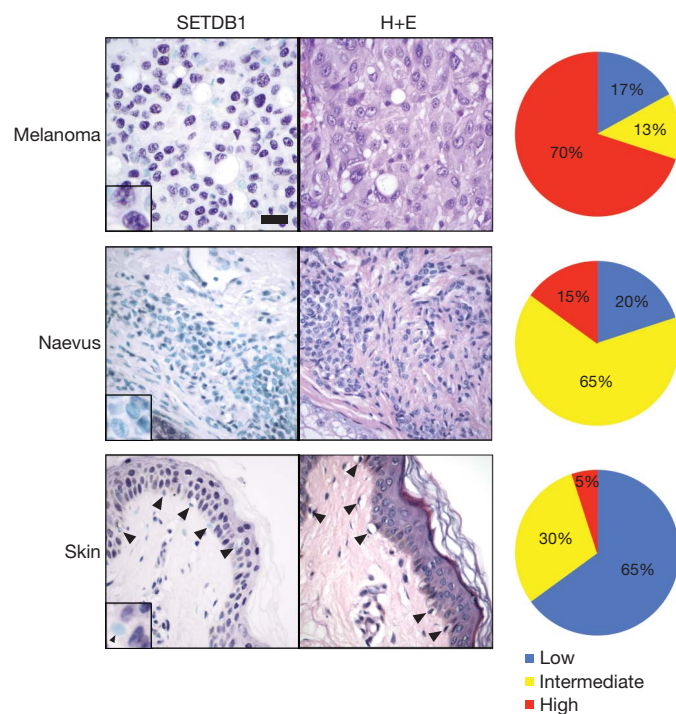
To obtain a mechanistic insight into the role of SETDB1 in regulating gene expression, we undertook genetic and biochemical studies that evaluate methyltransferase activity. Recently, a complex containing SETDB1 and the H3K9 methyltransferases SUV39H1, G9A (also known as EHMT2) and GLP (also known as EHMT1) was discovered<sup>16</sup>. To examine the possibility that other methyltransferases act together with SETDB1 to modulate melanoma onset, we tested whether



*SUV39H1* could accelerate melanoma formation in zebrafish. As was the case for *SETDB1*, overexpression of *SUV39H1* led to the formation of confluent melanocyte patches, and it accelerated melanoma onset (Fig. 3d). We next examined the consequences of mutations that render *SETDB1* enzymatically inactive. Enzymatically deficient *SETDB1* was capable of incorporating into the methyltransferase complex *in vitro* (Fig. 3e) and *in vivo* (Supplementary Fig. 15). Furthermore, in the context of enzymatically deficient *SETDB1*, the complex retained methyltransferase activity (Fig. 3f and Supplementary Fig. 16) and binding site localization (Supplementary Fig. 17). Last, the melanoma incidence curves for two methyltransferase-deficient *SETDB1* mutants were similar to each other and to the melanoma incidence curve for zebrafish that overexpress wild-type *SETDB1* (Fig. 3d). Our studies suggest a model in which activity of the methyltransferase complex containing *SETDB1* and *SUV39H1* alters gene expression in a way that leads to the acceleration of melanoma onset and to increased invasiveness.

To determine the extent of *SETDB1* overexpression in human melanomas, and to examine potential clinical implications, we performed immunohistochemistry on melanoma tissue microarrays. After confirming antibody specificity (Supplementary Fig. 18), we observed high levels of *SETDB1* expression in 5% of normal melanocytes ( $n = 20$ ), 15% of benign naevi ( $n = 20$ ) and 70% of malignant melanomas ( $n = 91$ ) (Fig. 4). On the basis of our observations of pre-malignant melanocytic lesions in zebrafish, we speculate that human naevi that overexpress *SETDB1* are more likely to undergo oncogenic progression than naevi with basal levels of *SETDB1* expression. These data indicate that the majority of malignant human melanomas overexpress the *SETDB1* protein.

In this study, we adapted the zebrafish as a platform for cancer gene discovery. Through the creation and analysis of more than 3,000



**Figure 4 | High level expression of *SETDB1* protein is common in human melanomas but not naevi or normal melanocytes.** Immunohistochemical staining of *SETDB1* (left, purple) and haematoxylin and eosin (H+E) staining (centre). Arrowheads indicate melanocytes in normal skin samples. *SETDB1* expression (right, measured as described in the Methods) was scored for malignant melanomas (top,  $n = 91$ ), naevi (centre,  $n = 20$ ) and normal skin (bottom,  $n = 20$ ). The percentage of samples with a low, intermediate or high level of *SETDB1* staining is indicated. Summarized data and raw data from experiments with two different antibodies are described in Supplementary Tables 4 and 5, respectively. Scale bar, 30  $\mu$ m; insets are magnified  $\times 2.5$ .

transgenic animals, *SETDB1* was identified as a gene capable of accelerating melanoma formation in cooperation with *BRAF(V600E)*. Amplification of the 1q21 chromosomal interval in melanoma does not preferentially occur together with the *BRAF(V600E)* mutation ( $P = 0.28$ , two-sample  $t$ -test). Therefore, it is probable that the tumour-promoting activity of *SETDB1* does not exclusively depend on *BRAF(V600E)*, which is common in melanomas but is found less frequently in other tumour types that have 1q21 amplification. *SETDB1* forms a multimeric complex with *SUV39H1* and other H3K9 methyltransferases. On the basis of our findings, we speculate that *SETDB1* overexpression can increase the activity of the H3K9 methyltransferase complex, leading to alterations in its target specificity. Inactivating mutations in histone methyltransferases and histone demethylases were recently described in renal cell carcinoma<sup>17,18</sup>. Our study lends functional support to the idea that perturbation of histone methylation promotes cancer. Moreover, *SETDB1* is focally amplified in a broad range of malignancies, suggesting that alterations in histone methyltransferase activity could define a biologically related subset of cancers.

## METHODS SUMMARY

**miniCoopR assay.** The miniCoopR vector was constructed by inserting a zebrafish *mitfa* minigene (consisting of promoter, open reading frame and 3'-untranslated region) into the BglII restriction site of the plasmid pDestTol2pA2 (ref. 19). Individual miniCoopR clones were created by MultiSite Gateway recombination (Invitrogen) using human full-length open reading frames. Twenty-five picograms of each miniCoopR-candidate clone and 25 pg mRNA encoding the Tol2 transposase were microinjected into one-cell zebrafish embryos generated from an incross of *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>; mitfa<sup>-/-</sup>* zebrafish. Rescued animals were scored weekly for the presence of visible tumours.

**Senescence assay.** SA- $\beta$ -Gal staining was performed as described previously<sup>10</sup>, except that scales plucked from the dorsum of melanocyte-rescued zebrafish were stained instead of tissue sections. This assay was performed on an *albino(b4)* mutant background so that melanin pigment would not obscure SA- $\beta$ -Gal staining. Experimental animals were injected with 20 pg miniCoopR-*SETDB1* plus 10 pg miniCoopR-EGFP, and control animals were injected with 30 pg miniCoopR-EGFP. Rescued melanocytes were recognized as EGFP-positive cells.

**Gene expression.** From zebrafish, total RNA was extracted from four miniCoopR-*SETDB1* melanomas and four miniCoopR-EGFP melanomas. Total RNA from each was used for the synthesis of cDNA, which was hybridized to a 385K microarray (NimbleGen 071105\_Zv7\_EXPR). Zebrafish genes that were downregulated by *SETDB1* were selected by a fold change of  $>5$  (when comparing the level of expression in miniCoopR-EGFP melanomas and the level in miniCoopR-*SETDB1* melanomas) and then filtered by a '*SETDB1* specificity score', which was defined as a fold change of  $>3$  when comparing the level of expression in *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>* melanomas with that of miniCoopR-*SETDB1* melanomas.

**Immunohistochemistry.** Human melanoma tissue microarrays were independently analysed for *SETDB1* protein by immunohistochemistry, using a rabbit polyclonal antibody (Sigma HPA018142, at a 1/200 dilution) and a mouse monoclonal antibody (4A3, Sigma WH0009869M7, 1/400 dilution), with a purple substrate for the secondary antibody (VIP substrate, Vector Labs). A methyl green counterstain was used.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 5 April 2010; accepted 7 January 2011.

- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Wan, P. T. *et al.* Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of *B-RAF*. *Cell* **116**, 855–867 (2004).
- Pollock, P. M. *et al.* High frequency of *BRAF* mutations in nevi. *Nature Genet.* **33**, 19–20 (2003).
- Curtin, J. A. *et al.* Distinct sets of genetic alterations in melanoma. *N. Engl. J. Med.* **353**, 2135–2147 (2005).
- Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
- Lin, W. M. *et al.* Modeling genomic diversity and tumor dependency in malignant melanoma. *Cancer Res.* **68**, 664–673 (2008).
- Patton, E. E. *et al.* *BRAF* mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **15**, 249–254 (2005).

8. Beroukhi, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
9. Michaloglou, C. *et al.* BRAF<sup>F600</sup>-associated senescence-like cell cycle arrest of human naevi. *Nature* **436**, 720–724 (2005).
10. Dimri, G. P. *et al.* A biomarker that identifies senescent human cells in culture and in aging skin *in vivo*. *Proc. Natl Acad. Sci. USA* **92**, 9363–9367 (1995).
11. Santoriello, C. *et al.* Expression of H-RASV12 in a zebrafish model of Costello syndrome causes cellular senescence in adult proliferating cells. *Dis. Model. Mech.* **2**, 56–67 (2009).
12. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
13. Mootha, V. K. *et al.* PGC-1 $\alpha$  responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34**, 267–273 (2003).
14. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23**, 2484–2489 (2009).
15. Wang, H. *et al.* mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol. Cell* **12**, 475–487 (2003).
16. Fritsch, L. *et al.* A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. *Mol. Cell* **37**, 46–56 (2010).
17. Dalglish, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
18. van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41**, 521–523 (2009).
19. Kwan, K. M. *et al.* The Tol2kit: a multisite gateway-based construction kit for *Tol2* transposon transgenesis constructs. *Dev. Dyn.* **236**, 3088–3099 (2007).
20. Beroukhi, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104**, 20007–20012 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Harrington, R. White and Y. Zhou for discussions; C. Lawrence, I. Adatto and L.-K. Zhang for expert fish care; G. Frampton for bioinformatics assistance; and K. Kwan, C.-B. Chien, and J. Boehm for reagents. This work was supported by grants from the Damon Runyon Cancer Research Foundation (C.J.C., DRG-1855-05), the Charles A. King Trust Foundation (C.J.C.), a Young Investigator Award from the American Society of Clinical Oncology (Y.H.), the Canadian Institutes of Health Research (S.B.) and the National Institutes of Health (C.J.C., K99AR056899-02; Y.H., K08DK075432-04; R.A.Y., CA146455, HG002668; and L.I.Z., CA103846 and DK055381).

**Author Contributions** C.J.C. and Y.H. contributed equally to this work and are listed alphabetically. C.J.C., Y.H. and L.I.Z. conceived the project, designed and analysed the experiments, and wrote the manuscript. C.J.C. and Y.H. performed the zebrafish experiments and contributed to the other experiments. J.J.-V. performed the tissue culture experiments. S.B. performed the ChIP-seq experiments and analysed the data. V.B., L.F., S.A.-S.-A. performed the biochemistry studies on SETDB1. L.A.J. performed the fluorescence *in situ* hybridization studies. T.J.H. performed the immunohistochemistry experiments. W.M.L., R.B. and C.H.M. analysed the copy number data. D.A.O. analysed the *SETDB1*-overexpression microarray data for WM451Lu cells. F.F. designed a database to manage and analyse tumour incidence data. C.B., C.J.B., L.T. and A.U. provided technical assistance. M.L., L.A.G. and R.A.Y. provided input into the preparation of the manuscript.

**Author Information** The data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus database under accession number GSE26372. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.I.Z. ([zon@enders.tch.harvard.edu](mailto:zon@enders.tch.harvard.edu)).

## METHODS

**miniCoopR assay.** The miniCoopR vector was constructed by inserting a zebrafish *mitfa* minigene (consisting of promoter, open reading frame and 3'-untranslated region) into the BglII site of the plasmid pDestTol2pA2 (ref. 19). Individual miniCoopR clones were created by MultiSite Gateway recombination (Invitrogen) using human full-length open reading frames. Recombination junctions were sequence verified. Twenty-five picograms of each miniCoopR-candidate clone and 25 pg mRNA encoding the Tol2 transposase were microinjected into one-cell zebrafish embryos generated from an incross of *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>; mitfa<sup>-/-</sup>* zebrafish. Transgenic animals were selected based on the presence of rescued melanocytes at 48 h post fertilization. Rescued animals were scored weekly for the presence of visible tumours.

**Tumour invasion assay.** Zebrafish with dorsal melanomas between the head and dorsal fin were isolated, and tumours were allowed to progress for 2 weeks, at which time the animals were killed. Tumours were formalin fixed, embedded and sectioned transversely to assess invasion.

**Senescence assay.** SA- $\beta$ -Gal staining was performed as described previously<sup>10</sup>, except that scales plucked from the dorsum of melanocyte-rescued zebrafish were stained instead of tissue sections. This assay was performed on an *albino(b4)* mutant background so that melanin pigment would not obscure SA- $\beta$ -Gal staining. Experimental animals were injected with 20 pg miniCoopR-*SETDB1* plus 10 pg miniCoopR-*EGFP*, and control animals were injected with 30 pg miniCoopR-*EGFP*. Rescued melanocytes were recognized as EGFP-positive cells.

**Gene expression and GSEA.** From zebrafish, total RNA was extracted from four miniCoopR-*SETDB1* melanomas and four miniCoopR-*EGFP* melanomas. Total RNA from each was used for the synthesis of cDNA, which was hybridized to a 385K microarray (NimbleGen 071105\_Zv7\_EXPR). Zebrafish genes that were downregulated by *SETDB1* were selected by a fold change of  $>5$  (when comparing the level of expression in miniCoopR-*EGFP* melanomas and the level in miniCoopR-*SETDB1* melanomas) and then filtered by a '*SETDB1* specificity score', which was defined as a fold change of  $>3$  when comparing the level of expression in *Tg(mitfa:BRAF(V600E)); p53<sup>-/-</sup>* melanomas with that of miniCoopR-*SETDB1* melanomas. Human orthologues of *SETDB1*-downregulated genes were identified for GSEA (<http://www.broadinstitute.org/gsea/>). For GSEA of *SETDB1*-downregulated and *SETDB1* 'bound-bound' genes, a rank-ordered gene list was derived from the expression profiles of 93 melanoma cell lines and short-term cultures of melanoma cells<sup>6</sup>, using *SETDB1* expression level as a continuous variable. In WM451Lu cells, the dose of *SETDB1* lentiviral infection was titrated to achieve *SETDB1* expression levels comparable to those in short-term cultures of melanoma cells with high levels of *SETDB1* expression. Total RNA was extracted and then amplified and hybridized to a Human Gene 1.0 ST Array (Affymetrix). Control gene expression values were obtained from WM451Lu cells infected with *EGFP*-expressing lentivirus.

**ChIP.** ChIP was performed from short-term cultures of WM262 and WM451Lu cells, and ChIP-seq data were analysed as described previously<sup>14</sup>.

**Methyltransferase complex reconstitution.** *In vitro*-translated Flag-GLP, HA-G9A and untagged SETDB1 (wild type (WT) or the indicated mutant) were incubated for 4 h at 4 °C with 5 mg GST, GST-SUV39H1(WT) or GST-SUV39H1(H324K) mutant immobilized on agarose-glutathione beads. Beads were then extensively washed, as described previously<sup>16</sup>, and protein complexes were eluted with free glutathione. The eluate was then subjected to an overnight Flag immunoprecipitation at 4 °C using Flag-agarose. After extensive washing, protein complexes were eluted with 0.1 M glycine, pH 3.0. The glycine was then neutralized with NaOH, and the eluate was renatured for 1 h at room temperature then incubated overnight at 4 °C with HA-resin. The HA-resin was then washed, and the protein complexes were eluted with SDS. Ten per cent of the input and 100% of the HA eluate were resolved by SDS-PAGE and were analysed by western blotting with the indicated antibodies. The top of the membrane was revealed with three different antibodies (anti-SETDB1, anti-HA and anti-Flag antibody) using two stripping steps.

**Histone methylation assay.** Purified complexes were incubated with 5 mg core histones (Upstate 13-107) and 1.5 mCi S-adenosyl-L-[methyl-<sup>3</sup>H]methionine (PerkinElmer NET155050UC) in a buffer containing 50 mM Tris, pH 8.0, 100 mM NaCl, 1% NP40, 1 mM dithiothreitol and protease inhibitors (with a reaction volume of 30  $\mu$ l). The mixture was incubated for 1 h at 30 °C and was then separated by SDS-PAGE. The gel was stained with a SimplyBlue SafeStain kit

(Invitrogen) and analysed by fluorography using an FLA-7000 phosphorimager (Fuji).

**Immunohistochemistry.** Human melanoma tissue microarrays were independently analysed for SETDB1 protein by immunohistochemistry, using a rabbit polyclonal antibody (Sigma HPA018142, at a 1/200 dilution) and a mouse monoclonal antibody (4A3, Sigma WH0009869M7, 1/400 dilution), with a purple substrate for the secondary antibody (VIP substrate, Vector Labs). A methyl green counterstain was used. Melanoma tissue microarrays were obtained from US Biomax (ME1003 and ME482). A modified visual semiquantitative method was used to score staining as described previously<sup>21</sup>, using a two-score system for immunointensity (II) and immunopositivity (IP). II and IP were then multiplied. SETDB1 immunostaining was also performed on formalin-fixed, paraffin-embedded zebrafish melanomas.

**Fluorescence *in situ* hybridization (FISH).** The bacterial artificial chromosome (BAC) clone, RP11-42A12, used as the *SETDB1* probe was selected using the UCSC Genome Browser and obtained from the BACPAC Resource Center (CHORI). BAC probe preparation, labelling and hybridization were performed as described previously<sup>22</sup>. The SpectrumOrange-CEP1 reference probe 06J39-026 was obtained from Abbott Molecular.

**Lentivirus infection.** We used pLKO1-puromycin lentiviral vectors carrying short hairpin RNAs (shRNAs) specific for *SETDB1* or *GFP* sequences. The shRNA vectors were obtained from the Broad Institute RNAi Consortium (<http://www.broadinstitute.org/rnai/trc>), and the lentiviruses were obtained by overnight triple co-transfection of 293T cells using lipofectamine 2000 and 3  $\mu$ g pLKO.1-shRNA,  $\Delta$ 8.9 and VSV-G vectors (in 100-mm plates). The *SETDB1* shRNA construct used was TRCN0000148112 (hairpin target sequence gtcagat gataactctgta). Knockdown efficiency was determined by RT-PCR and by immunoblot analysis using *SETDB1*-specific primers and antibody, respectively. At days 2 and 3 post infection, supernatants were collected and filtered with a 45- $\mu$ m filter to remove 293T cells. Virus was added to cells (plated to attain 30–40% confluence at the time of infection) and incubated for 16 h in the presence of polybrene at 5  $\mu$ g ml<sup>-1</sup>. After infection, virus was removed and fresh media added. Forty-eight hours post infection, cells were subjected to a 3-day puromycin selection. To elevate *SETDB1* expression in WM451Lu cells, we used a pLEX980 lentiviral vector into which the wild-type *SETDB1* open reading frame had been recombined. Infection, selection and monitoring of SETDB1 concentration were performed as described above.

**Cell proliferation assays.** Cells were plated in 12-well plates at 20,000 cells well<sup>-1</sup> in 2 ml medium. At each time point, cells were trypsinized from 3 wells and counted with a cell counter (Beckman Coulter).

**Western blot analyses.** Western blots were performed with primary antibodies recognizing SETDB1 (Abcam ab12317), BRAF (Santa Cruz Biotechnology sc5284),  $\alpha$ -tubulin (Cell Signaling Technology 2144), SUV39H1 (Upstate 07-550), the Flag epitope (Sigma M2), G9A (Clinisciences D141-3), and GLP (R&D Systems PP-B0422-00). HRP-conjugated anti-rabbit and anti-mouse secondary antibodies (Amersham) were used. Thirty micrograms total protein was loaded per lane. SETDB1 was recognized as a doublet at approximately 150 kDa (predicted molecular mass 143 kDa).

**p53BP2 quantitative PCR analysis.** Chromatin extracts from HeLa cells were prepared as described<sup>16</sup>, using micrococcal nuclease digestion and mild sonication (without any formaldehyde crosslinking) to enrich them in mono-nucleosomes. Flag-HA purification was then performed. A fraction of the input or 1/3 of each Flag-HA purified complex (chromatin associated) was treated with proteinase K and RNase. Samples were then phenol-chloroform extracted, and DNA was precipitated using ethanol. p53BP2 primers were described previously<sup>23</sup>. Results were normalized to input values and presented as fold enrichment compared with the control sample (Flag-HA purification from HeLa control cells).

- Nisolle, M. *et al.* Immunohistochemical study of the proliferation index, oestrogen receptors and progesterone receptors A and B in leiomyomata and normal myometrium during the menstrual cycle and under gonadotrophin-releasing hormone agonist therapy. *Hum. Reprod.* **14**, 2844–2850 (1999).
- Perner, S. *et al.* EML4-ALK fusion lung cancer: a rare acquired event. *Neoplasia* **10**, 298–302 (2008).
- Sarraf, S. A. & Stancheva, I. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15**, 595–605 (2004).



# DHODH modulates transcriptional elongation in the neural crest and melanoma

Richard Mark White<sup>1,2</sup>, Jennifer Cech<sup>1</sup>, Sutteera Ratanasirintra<sup>1</sup>, Charles Y. Lin<sup>3,4</sup>, Peter B. Rahl<sup>3</sup>, Christopher J. Burke<sup>1</sup>, Erin Langdon<sup>1</sup>, Matthew L. Tomlinson<sup>5</sup>, Jack Mosher<sup>6</sup>, Charles Kaufman<sup>1,2</sup>, Frank Chen<sup>7</sup>, Hannah K. Long<sup>8</sup>, Martin Kramer<sup>9</sup>, Sumon Datta<sup>1</sup>, Donna Neuberg<sup>10</sup>, Scott Granter<sup>11</sup>, Richard A. Young<sup>3,4</sup>, Sean Morrison<sup>6</sup>, Grant N. Wheeler<sup>3</sup> & Leonard I. Zon<sup>1</sup>

Melanoma is a tumour of transformed melanocytes, which are originally derived from the embryonic neural crest. It is unknown to what extent the programs that regulate neural crest development interact with mutations in the *BRAF* oncogene, which is the most commonly mutated gene in human melanoma<sup>1</sup>. We have used zebrafish embryos to identify the initiating transcriptional events that occur on activation of human *BRAF(V600E)* (which encodes an amino acid substitution mutant of BRAF) in the neural crest lineage. Zebrafish embryos that are transgenic for *mitfa:BRAF(V600E)* and lack *p53* (also known as *tp53*) have a gene signature that is enriched for markers of multipotent neural crest cells, and neural crest progenitors from these embryos fail to terminally differentiate. To determine whether these early transcriptional events are important for melanoma pathogenesis, we performed a chemical genetic screen to identify small-molecule suppressors of the neural crest lineage, which were then tested for their effects on melanoma. One class of compound, inhibitors of dihydroorotate dehydrogenase (DHODH), for example leflunomide, led to an almost complete abrogation of neural crest development in zebrafish and to a reduction in the self-renewal of mammalian neural crest stem cells. Leflunomide exerts these effects by inhibiting the transcriptional elongation of genes that are required for neural crest development and melanoma growth. When used alone or in combination with a specific inhibitor of the *BRAF(V600E)* oncogene, DHODH inhibition led to a marked decrease in melanoma growth both *in vitro* and in mouse xenograft studies. Taken together, these studies highlight developmental pathways in neural crest cells that have a direct bearing on melanoma formation.

In melanoma, it is unknown to what extent *BRAF(V600E)* mutations depend on transcriptional programs that are present in the developmental lineage of tumour initiation. These programs may be therapeutic targets when combined with *BRAF(V600E)* inhibition. We have used zebrafish embryos to identify small-molecule suppressors of neural crest progenitors that give rise to melanoma. Transgenic zebrafish expressing human *BRAF(V600E)* under the melanocyte-specific *mitfa* promoter, *Tg(mitfa:BRAF(V600E))*, develop melanoma at 4–12 months of age when crossed with *p53*<sup>-/-</sup> mutant zebrafish, *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> (Fig. 1a). Because the *mitfa* promoter drives the expression of *BRAF(V600E)* from 16 h after fertilization (a time point that overlaps with the expression of embryonic neural crest markers such as *sox10*), events that occur early in embryogenesis are analogous to those that occur at tumour initiation. To gain insight into these initiating events, we compared the gene expression profiles of *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> embryos with those of *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> melanomas by using gene set enrichment analysis (Fig. 1b). This approach uncovered a signature of 123

overlapping genes, which is enriched for markers of embryonic neural crest progenitors (*crestin*, *sox10* and *ednrb* (also known as *ednrb1*)) and melanocytes (*tyr* and *dct*) (see Supplementary Table 1 for full gene sets). The overlapping gene signature is similar to the signature of a multipotent neural crest progenitor, suggesting that the melanomas have adopted this cell fate.

We analysed alterations in embryonic neural crest development by using *in situ* hybridization. At 24 h post fertilization, *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> embryos show an abnormal expansion in the number of *crestin*<sup>+</sup> progenitors, together with an increase in other markers from the 123 gene signature such as *spry4* and *rab31l1* (Supplementary Fig. 1). By 72 h post fertilization, *crestin* persists aberrantly in the head, tail and dorsal epidermis only in *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> embryos but not in embryos with either single mutation (Supplementary Fig. 2a). The gene encoding *Crestin* is zebrafish specific<sup>2</sup> and is normally downregulated after the terminal differentiation of neural crest progenitors<sup>3</sup>. Our finding therefore suggests that activated *BRAF(V600E)* promotes the maintenance of multipotency in neural crest progenitors, which become expanded during tumorigenesis. In adult *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> melanomas, almost all tumour cells, but no normal cells, were positive for *crestin* (Fig. 1c). Only 10–15% of the melanoma cells were pigmented (Supplementary Fig. 2b), which is consistent with the concept that adult zebrafish melanomas retain a progenitor-like state. A human melanoma tissue array yielded similar results: 52 out of 70 of the melanomas on the array (74.3%) were positive for the neural crest progenitor gene *ednrb*, but only 9 of 70 (12.9%) were positive for the melanocyte lineage marker *dct* (Supplementary Fig. 3), in agreement with the finding that most human melanomas express the neural crest marker *sox10* (ref. 4). These data indicate that the majority of human melanomas reflect events that lead to the maintenance of a neural crest progenitor phenotype<sup>5</sup>.

We proposed that chemical suppressors of neural crest progenitors would be useful for treating melanoma. We screened 2,000 chemicals to identify compounds that inhibit the *crestin*<sup>+</sup> lineage during embryogenesis. Most chemicals (90%) had a minimal effect or were toxic (Supplementary Fig. 4). NSC210627, a molecule of unknown function, strongly abrogated the expression of *crestin* (Fig. 2a, centre and left). The chemoinformatic algorithm DiscoveryGate<sup>6</sup> revealed similarity between NSC210627 and brequinar (Supplementary Fig. 5), an inhibitor of DHODH<sup>7</sup>. NSC210627 inhibited DHODH activity *in vitro* (Supplementary Fig. 6). Leflunomide, a DHODH inhibitor that is structurally distinct from NSC210627 (ref. 8), phenocopied NSC210627 (Fig. 2a, right) and was used for further studies because of its availability.

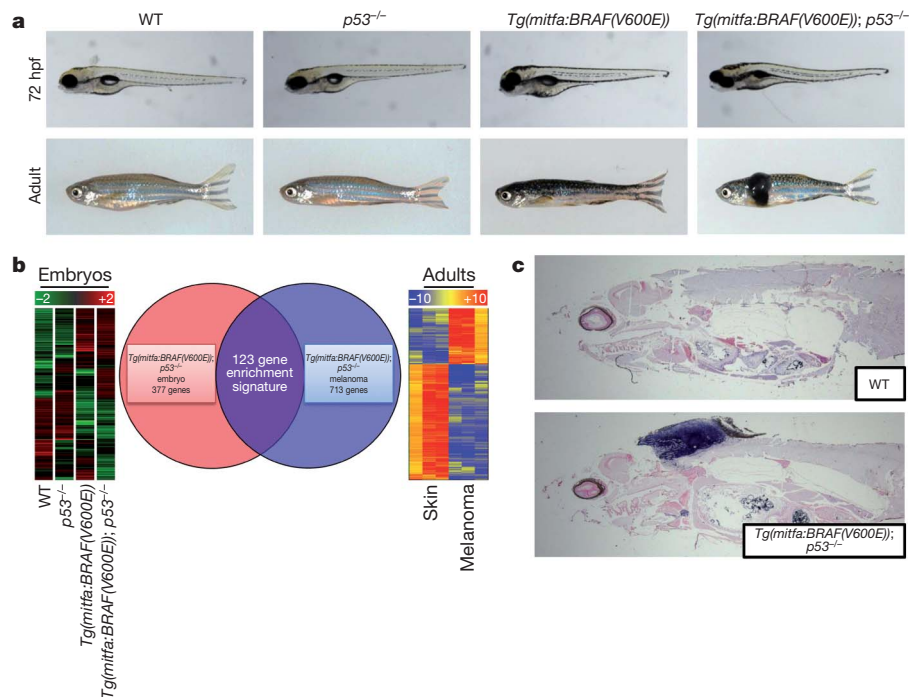
We examined neural crest derivatives affected by leflunomide. Treated zebrafish embryos were devoid of pigmented melanocytes at 36–48 h post fertilization (Fig. 2b) and iridophores at 72 h post fertilization

<sup>1</sup>Stem Cell Program and Hematology/Oncology, Children's Hospital Boston, Howard Hughes Medical Institute, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02115, USA.

<sup>2</sup>Department of Medical Oncology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. <sup>3</sup>Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA. <sup>4</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. <sup>5</sup>School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK. <sup>6</sup>Center for Stem Cell Biology, University of Michigan, Howard Hughes Medical Institute, Ann Arbor, Michigan 48109, USA. <sup>7</sup>Harvard University, Cambridge, Massachusetts 02138, USA. <sup>8</sup>Cambridge University, Cambridge CB2 1TN, UK.

<sup>9</sup>Genzyme Corporation, Cambridge, Massachusetts 02142, USA. <sup>10</sup>Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA.

<sup>11</sup>Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.



**Figure 1 | Transgenic zebrafish melanoma and neural crest gene expression.**

**a**, Transgenic zebrafish expressing BRAF(V600E) under the control of the promoter of the melanocyte-specific gene *mitfa*,  $Tg(mitfa:BRAF(V600E))$ , develop pigmentation abnormalities and melanoma when crossed with  $p53^{-/-}$  fish. Their gross embryonic development is largely normal. hpf, hours post fertilization; WT, wild type. **b**, Gene expression analysis revealed a unique gene signature at 72 h post fertilization in the  $Tg(mitfa:BRAF(V600E)); p53^{-/-}$  strain (left). Gene set enrichment analysis uncovered an enrichment between the

(Supplementary Fig. 7a). *dhodh* inhibition led to a loss of ventral melanocytes in stage 38 *Xenopus* embryos (Supplementary Fig. 7b). Leflunomide treatment led to an almost complete loss of melanocyte progenitors at 24 h post fertilization (Fig. 2c), a reduction in the number of glial cells at 72 h post fertilization (Fig. 2d) and disruption of jaw cartilage at 72 h post fertilization (data not shown). Leflunomide reduced the expression of *sox10* and *dct*, which are expressed by neural crest progenitors and melanocytes, respectively, while leaving other lineages such as blood and notochord less affected (Supplementary Fig. 8). Microarray analysis of leflunomide-treated embryos showed downregulation of 49% of the genes that were upregulated in the 123-gene melanoma signature, and more than half of these are neural crest related (see Supplementary Table 2 for the complete list).

The loss of several types of neural crest derivative suggested that leflunomide acts on neural crest stem cells. We tested leflunomide, and its derivative A771726, on neural crest stem cells isolated from the fetal (embryonic day 14.5) rat gut<sup>9,10</sup>. Both chemicals reduced the number of self-renewing neural crest stem cells in primary stem cell colonies, to  $27 \pm 5.35\%$  (leflunomide) and  $35 \pm 6.16\%$  (A771726) of control numbers ( $P < 0.0003$  and  $P < 0.00007$ , respectively, Student's *t*-test; Fig. 2e and Supplementary Fig. 9a). Colony size was also reduced compared with controls (by 18% and 24%, respectively;  $P < 0.02$ , Student's *t*-test), but there was no effect on the differentiation or survival of specific progeny (Supplementary Fig. 9b, c). These results demonstrate that DHODH inhibitors negatively regulate the self-renewal of neural crest stem cells and have an effect on these cells in multiple species.

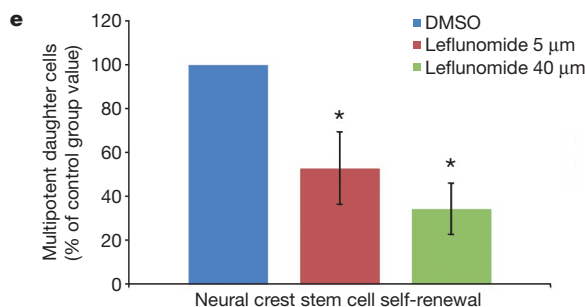
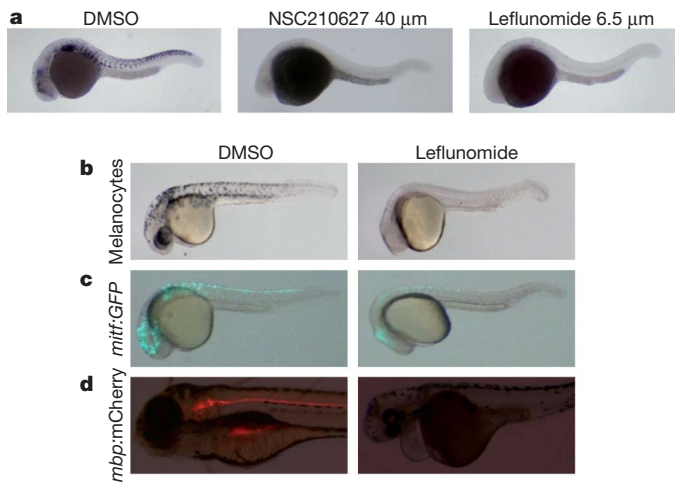
DHODH catalyses the fourth step in the synthesis of pyrimidine nucleotides (NTPs)<sup>11</sup>. We noted striking morphological similarity between leflunomide-treated embryos and *spt5/spt6* mutants<sup>12</sup>, suggesting that leflunomide acts to suppress transcriptional elongation. In the *spt5<sup>sk8</sup>* null mutant, we found a lack of both *crestin* expression and pigmented melanocytes (similar to leflunomide-treated embryos) (Supplementary Fig. 10a). At 24 h post fertilization, the gene expression

embryonic gene signature and the adult melanomas that form 4–12 months later (centre and right; see the Supplementary Information for full protocol details). Embryo heat-map columns represent an average of three clutches (log<sub>2</sub> scale, range –2-fold to +2-fold increase); adult heat-map columns represent individual fish (log<sub>2</sub> scale, range –10-fold to +10-fold increase). **c**, *In situ* hybridization of sagittal sections of WT and  $Tg(mitfa:BRAF(V600E)); p53^{-/-}$  adults reveal homogeneous *crestin* expression (blue) only within the dorsal melanoma; it is absent from normal adult tissues.

profiles of *spt5<sup>sk8</sup>* mutants and leflunomide-treated embryos<sup>13</sup> were nearly identical; of 223 genes downregulated after leflunomide treatment, 183 were similarly downregulated in *spt5<sup>sk8</sup>* mutants (Supplementary Table 3 and Supplementary Fig. 10b). These downregulated genes include neural crest genes (*crestin*, *sox10* and *mitfa*) and members of the *notch* pathway (*her2* and *dllb*). We examined the interaction of Dhodh with *spt5* by incubating the hypomorphic *spt5<sup>m806</sup>* mutant (which has only mild melanocyte defects)<sup>14</sup> in low concentrations of leflunomide (3–5  $\mu$ M) and then analysing the number of pigmented melanocytes. Enhanced sensitivity to leflunomide was shown by *spt5<sup>m806</sup>* embryos (Fig. 3a and Supplementary Fig. 11); at 3  $\mu$ M leflunomide, 99% of mutant embryos had few or no melanocytes, compared with 0% of wild-type embryos ( $P = 0.000018$ , Kruskal–Wallis test; Supplementary Fig. 11b). These data confirm that DHODH inhibition affects transcriptional elongation, which is consistent with previous data demonstrating that a reduction in nucleotide pools *in vitro* leads to defects in elongation<sup>15</sup>.

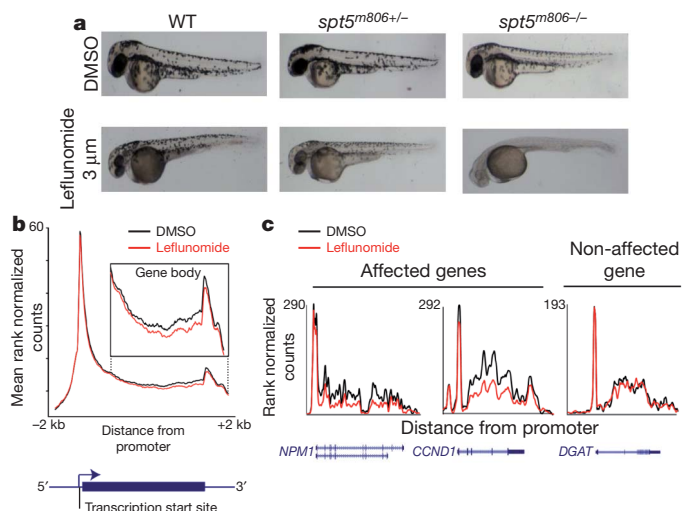
We assessed whether leflunomide specifically caused defects in the transcriptional elongation of genes required for neural crest development by using reverse transcription followed by quantitative PCR (Supplementary Fig. 10c and Supplementary Table 4). Leflunomide caused either no change or an increase in 5' transcript abundance but a significant downregulation of 3' transcripts of *mitfa* (for 5' transcripts  $3.75 \pm 1.19$ -fold increase versus  $0.39 \pm 0.07$ -fold increase for 3' transcripts;  $P < 0.05$ , Student's *t*-test) and *dllb* (5' transcripts  $1.13 \pm 0.14$ -fold increase versus 3' transcripts  $0.74 \pm 0.07$ -fold increase;  $P < 0.05$ ). Leflunomide did not have a similar effect on control genes such as the gene encoding  $\beta$ -actin (5' transcripts  $1.03 \pm 0.07$ -fold increase versus 3' transcripts  $0.99 \pm 0.06$ -fold increase; *P* is not significant, Student's *t*-test). In the presence of leflunomide, transcription is initiated normally, but the transcripts accumulate and do not undergo productive elongation.

To confirm that this mechanism is conserved in human melanoma, we performed chromatin immunoprecipitation using an antibody



**Figure 2 | A chemical genetic screen to identify suppressors of neural crest development.** **a**, A search for chemical suppressors of the *crestin*<sup>+</sup> lineage during embryogenesis identified NSC210627, a compound that completely abrogates the expression of *crestin* (which is normally present in the head, along the dorsum and in ventrally migrating neural crest cells), as shown by *in situ* hybridization (**a**, left and centre). DMSO is used as a control. The DiscoveryGate chemoinformatic algorithm revealed structural similarity between NSC210627 and brequinar (Supplementary Fig. 5), an inhibitor of DHODH. Leflunomide, a structurally distinct DHODH inhibitor, phenocopies the *crestin* phenotype induced by treatment with NSC210627 (**a**, right). **b–d**, Treatment with leflunomide caused an absence of multiple neural crest derivatives, including pigmented melanocytes (**b**); melanocyte progenitors, which were visualized by expressing green fluorescent protein (GFP) under the control of the *mitf* promoter (**c**); and glial cells, which were visualized by expressing the fluorescent protein mCherry under the control of the myelin basic protein (*mbp*) promoter (**d**). **e**, Treatment with leflunomide, or A771726 (Supplementary Fig. 9a), significantly reduced the number of multipotent daughter cells that could be subcloned from individual primary neural crest stem cell colonies. Values shown are mean  $\pm$  s.d. of three replicates; \*,  $P < 0.05$  compared with control, Student's *t*-test.

specific for RNA polymerase II (RNA Pol II), followed by sequencing (ChIP-seq). Transcriptional elongation was measured using the travelling ratio (TR)<sup>16</sup>, in which the ratio of RNA Pol II density in the promoter-proximal region is compared with that in the gene body. In the human melanoma cell lines A375 and Malme-3M, leflunomide treatment caused a significant inhibition of transcriptional elongation (measured as an increase in the TR), particularly for genes with a TR that was initially low ( $<7.5$ ). For example, in A375 cells, the TR increased by  $>1.3$  fold at 21.3% of loci; in Malme-3M cells, this occurred for 36.3% of loci (Supplementary Table 5). Examination of RNA Pol II occupancy using metagenome analysis at a variety of fold-change cutoffs (Fig. 3b (A375), Supplementary Fig. 12 (Malme-3M) and Supplementary Table 5) revealed no defect in transcription initiation but a decrease in elongation that was pronounced at the 3' end of genes such as *NPM1* and *CCND1* (Fig. 3c). Ingenuity Pathway Analysis on the loci affected in both cell lines revealed a strong enrichment for Myc targets and pathway members<sup>17</sup> (Supplementary Fig.



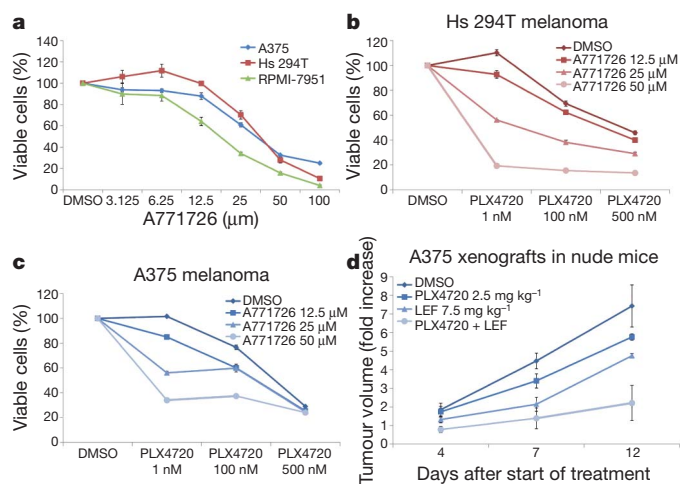
**Figure 3 | DHODH inhibition modulates transcriptional elongation.** **a**, The hypomorphic *spt5*<sup>m806</sup> homozygous mutant (top right) has only a mild pigmentation defect compared with WT (top left) or heterozygous (top centre) animals. Treatment with a low dose of leflunomide (3  $\mu$ M) leads to an almost complete absence of neural-crest-derived melanocytes in the mutant line. See Supplementary Fig. 11 for dose-response quantification of this effect. **b**, Metagenome analysis of RNA Pol II occupancy in A375 human melanoma cells after treatment with leflunomide. The metagenome plot allows visualization of all of the genes that are occupied by RNA Pol II, corrected for individual gene lengths. Genome-wide RNA Pol II occupancy at the promoter region is unaffected but is diminished at the 3' end of the genes. The inset shows a higher magnification of the 3' region of the genes. **c**, Representative examples of Myc target genes (*NPM1* and *CCND1*), which demonstrate defects in transcriptional elongation after treatment with leflunomide. A gene that is minimally affected (*DGAT*) is also depicted. For *NPM1*, the TR is 5.04 after DMSO treatment and 8.10 after leflunomide treatment. For *CCND1*, the TR is 3.47 after DMSO treatment and 4.67 after leflunomide treatment. For *DGAT*, the TR is 5.19 after DMSO treatment and 5.34 after leflunomide treatment.

13a, b). Myc, in addition to its requirement for neural crest development<sup>18</sup>, was recently shown to be a potent regulator of transcriptional pause release in embryonic stem cells<sup>16</sup>. Our data suggest that the mechanism by which Myc target genes are regulated at the transcriptional elongation level operates in neural-crest-derived melanoma as well. Taken together, the genetic and biochemical data demonstrate that leflunomide acts to modulate transcriptional elongation in both neural crest development and human melanoma.

Given the effect of DHODH inhibition on neural crest development, we tested its effects on melanoma growth. A771726 caused a dose-dependent decrease in the proliferation of human melanoma cell lines (A375, Hs 294T and RPMI-7951; Fig. 4a). Similarly, a short hairpin RNA directed against DHODH led to a 57.7% decrease in the proliferation of A375 cells, as well as a decrease in elongation as measured by ChIP-PCR (Supplementary Fig. 14). Microarray analysis of the A375 cell line treated with leflunomide revealed downregulation of genes that are required for neural crest development (such as *SNAI2*) and members of the *NOTCH* pathway (for example, *HES6* and *JAG1*), which is consistent with the effects of leflunomide on zebrafish embryos (Supplementary Table 6).

Pyrimidine NTP production is regulated at the level of carbamoyl-phosphate synthetase (CAD)<sup>19</sup>, the enzyme that is directly upstream of DHODH. CAD is phosphorylated by the mitogen-activated protein kinase ERK2<sup>20</sup>, a protein that would be activated in melanoma as a result of the BRAF(V600E) mutation. We reasoned that combined blockade of BRAF(V600E) and DHODH would suppress melanoma growth. We measured melanoma cell proliferation after treatment with the BRAF(V600E) inhibitor PLX4720 (ref. 21) together with A771726 (Fig. 4b, c and Supplementary Fig. 15a, b), and we found that the combination led to a cooperative suppression of melanoma growth.





**Figure 4 | Melanoma growth is suppressed by DHODH blockade in concert with BRAF(V600E) inhibition.** **a**, A771726 causes a dose-dependent decrease in melanoma cell proliferation, as measured by the CellTiter-Glo assay, in three human melanoma cell lines that contain the BRAF(V600E) mutation (A375, Hs 294T and RPMI-7951). **b**, **c**, A771726 cooperates with the BRAF(V600E) inhibitor PLX4720 in inhibiting melanoma cell proliferation in the Hs 294T (**b**) and A375 (**c**) cell lines, as well as other melanoma cell lines (Supplementary Fig. 15). **d**, After subcutaneous transplantation of  $3 \times 10^5$  A375 cells into nude mice ( $n = 4-5$ ), both leflunomide (LEF) alone and PLX4720 alone impair tumour progression; the combination of these chemicals elicits an almost complete abrogation of tumour growth and results in complete tumour regression in two of five animals (DMSO versus PLX4720,  $P = 0.036$ ; DMSO versus LEF,  $P = 0.006$ ; PLX4720 or LEF versus PLX4720 + LEF,  $P = 0.006$ ; PLX4720 versus LEF,  $P$  is not significant; analysis of variance, followed by Tukey's post hoc test). **a-d**, Values shown are mean  $\pm$  s.e.m. of three to five replicates.

PLX4720 had no effect in non-melanoma cell lines (Supplementary Fig. 15c, BRAF<sup>WT</sup>). A771726 demonstrated some antiproliferative activity in non-melanoma cell lines but was less potent in these cells than in melanoma cells (Supplementary Fig. 15d).

We examined the *in vivo* effects of leflunomide and PLX4720 by using xenografts of A375 cells transplanted into nude mice (Fig. 4c and Supplementary Fig. 16). At 12 days post treatment, tumours in mice that had been treated with dimethylsulphoxide (DMSO) as a control had grown  $7.4 \pm 1.3$  fold. By contrast, tumours in PLX4720-treated mice had grown  $5.7 \pm 0.16$ -fold, and those in leflunomide-treated mice had grown  $4.7 \pm 0.12$  fold. The combination of PLX4720 and leflunomide led to an enhanced abrogation of tumour growth, with only  $2.2 \pm 0.9$  fold growth. In 40% of animals, this combination led to almost complete tumour regression (PLX4720 and leflunomide versus PLX4720 alone or leflunomide alone,  $P < 0.001$ , analysis of variance followed by Tukey's post hoc test). Therefore, we have found that an inhibitor of embryonic neural crest development, leflunomide, blocks *in vivo* tumour growth in combination with the BRAF(V600E) inhibitor PLX4720 when used at clinically meaningful doses.

Our data suggest that inhibition of DHODH abrogates the transcriptional elongation of genes that are required for both neural crest development and melanoma growth, including Myc target genes and *mitfa*. Although DHODH inhibition would be expected to lead to ubiquitous defects, human mutations in *DHODH* cause Miller's syndrome<sup>22</sup>, a craniofacial disorder that is similar to syndromes with defects in neural crest development. Our data support recent findings that elongation factors are important for both neural crest development<sup>23</sup> and cancer growth<sup>24</sup>. Developmental regulators of transcriptional elongation have recently been identified to be important in haematopoiesis<sup>25</sup>, and identification of such factors in the neural crest awaits further study.

Using chemical genetic approaches in zebrafish and *Xenopus* allows the identification of molecules that require *in vivo* contexts for the

expression of relevant phenotypes<sup>26</sup>. Inhibition of DHODH may be a unique *in vivo* mechanism for modulating transcriptional elongation. Leflunomide is a well-tolerated anti-arthritis drug in humans<sup>27</sup>, and our data suggest that, for the treatment of melanoma, it would be most effective in combination with a BRAF(V600E) inhibitor. This combination therapy may help to overcome resistance to BRAF(V600E) inhibitors<sup>28</sup>. As an increasing number of genomic changes are identified in cancerous cells, the challenge is to target these in concert with lineage-specific factors to achieve therapeutic synergy. Our approach to identifying lineage-specific suppressors in zebrafish embryos can be generalized to other cell types, with direct relevance to human cancer.

## METHODS SUMMARY

Microarray analysis was performed on four groups of embryos at 72 h post fertilization: wild type, *Tg(mitfa:BRAF(V600E))*, *p53*<sup>-/-</sup>, and *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> double mutants. Array analysis was also performed for adult *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> melanomas and for adjacent skin. The transcriptional signature of the melanomas was used in gene set enrichment analysis to identify genes that were significantly enriched in the *Tg(mitfa:BRAF(V600E)); p53*<sup>-/-</sup> embryos. The 123 genes that make up this signature, which is enriched for markers of the neural crest, were concordantly upregulated or down-regulated in both *BRAF(V600E); p53*<sup>-/-</sup> embryos and tumours. *In situ* hybridization was used to examine the expression of *crestin* (a pan neural crest marker) and that of other neural crest genes in embryos (at 24–72 h post fertilization) and adult tumours. Chemical screening was performed to identify suppressors of the *crestin*<sup>+</sup> lineage by treating wild-type embryos from 50% epiboly to 24 h post fertilization with various chemicals, followed by robot-controlled *in situ* hybridization. Two inhibitors of DHODH abrogated *crestin* expression: NSC210627 and leflunomide. The latter was used for further studies owing to its more widespread availability. The effect of leflunomide on embryonic neural crest development in zebrafish was assessed by scoring embryos for melanocytes, iridophores and glial cells. Leflunomide was further assessed for its ability to affect the multipotent self-renewal of purified p75<sup>+</sup>α<sub>4</sub>-integrin<sup>+</sup> rat neural crest stem cells (referred to in the main text as neural crest stem cells). The effects of leflunomide on transcriptional elongation in the neural crest were tested using the *spts*<sup>m806</sup> allele and measuring pigmentation in response to 3–5 μM leflunomide. Elongation in melanoma cells was assayed by ChIP-seq using an antibody specific for RNA Pol II and measuring the TR. Leflunomide was tested for anti-melanoma effects on human melanoma cell lines in the presence or absence of the BRAF(V600E) inhibitor PLX4720. *In vitro* proliferation assays were performed using the CellTiter-Glo system (Promega). *In vivo* effects were tested by treating established A375 xenografts with daily intraperitoneal doses of PLX4720 alone, leflunomide alone or both, and the tumour growth rate was measured on days 4, 7 and 12.

Received 26 March 2010; accepted 31 January 2011.

- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- Rubinstein, A. L., Lee, D., Luo, R., Henion, P. D. & Halpern, M. E. Genes dependent on zebrafish *cyclops* function identified by AFLP differential gene expression screen. *Genesis* **26**, 86–97 (2000).
- Luo, R., An, M., Arduini, B. L. & Henion, P. D. Specific pan-neural crest expression of zebrafish *Crestin* throughout embryonic development. *Dev. Dyn.* **220**, 169–174 (2001).
- Bakos, R. M. *et al.* Nestin and SOX9 and SOX10 transcription factors are coexpressed in melanoma. *Exp. Dermatol.* **19**, e89–e94 (2010).
- Boiko, A. D. *et al.* Human melanoma-initiating cells express neural crest nerve growth factor receptor CD271. *Nature* **466**, 133–137 (2010).
- DiscoveryGate (<https://www.discoverygate.com/>) (2008).
- McLean, J. E., Neidhardt, E. A., Grossman, T. H. & Hedstrom, L. Multiple inhibitor analysis of the brequinar and leflunomide binding sites on human dihydroorotate dehydrogenase. *Biochemistry* **40**, 2194–2200 (2001).
- Kaplan, M. J. Leflunomide Aventis Pharma. *Curr. Opin. Invest. Drugs* **2**, 222–230 (2001).
- Bixby, S., Kruger, G. M., Mosher, J. T., Joseph, N. M. & Morrison, S. J. Cell-intrinsic differences between stem cells from different regions of the peripheral nervous system regulate the generation of neural diversity. *Neuron* **35**, 643–656 (2002).
- Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
- Löffler, M., Jockel, J., Schuster, G. & Becker, C. Dihydroorotate-ubiquinone oxidoreductase links mitochondria in the biosynthesis of pyrimidine nucleotides. *Mol. Cell. Biochem.* **174**, 125–129 (1997).
- Keegan, B. R. *et al.* The elongation factors Pandora/Spt6 and Foggy/Spt5 promote transcription in the zebrafish embryo. *Development* **129**, 1623–1632 (2002).

13. Krishnan, K., Salomonis, N. & Guo, S. Identification of Spt5 target genes in zebrafish development reveals its dual activity *in vivo*. *PLoS ONE* **3**, e3621 (2008).
14. Guo, S. *et al.* A regulator of transcriptional elongation controls vertebrate neuronal development. *Nature* **408**, 366–369 (2000).
15. Wada, T. *et al.* DSIF, a novel transcription elongation factor that regulates RNA polymerase II processivity, is composed of human Spt4 and Spt5 homologs. *Genes Dev.* **12**, 343–356 (1998).
16. Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
17. Chen, Y. *et al.* Integration of genome and chromatin structure with gene expression profiles to predict c-MYC recognition site binding and function. *PLoS Comput. Biol.* **3**, e63 (2007).
18. Hong, S. K., Tsang, M. & Dawid, I. B. The *mych* gene is required for neural crest survival during zebrafish development. *PLoS ONE* **3**, e2029 (2008).
19. Aoki, T., Morris, H. P. & Weber, G. Regulatory properties and behavior of activity of carbamoyl phosphate synthetase II (glutamine-hydrolyzing) in normal and proliferating tissues. *J. Biol. Chem.* **257**, 432–438 (1982).
20. Graves, L. M. *et al.* Regulation of carbamoyl phosphate synthetase by MAP kinase. *Nature* **403**, 328–332 (2000).
21. Tsai, J. *et al.* Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl Acad. Sci. USA* **105**, 3041–3046 (2008).
22. Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nature Genet.* **42**, 30–35 (2010).
23. Nguyen, C. T., Langenbacher, A., Hsieh, M. & Chen, J. N. The PAF1 complex component Leo1 is essential for cardiac and neural crest development in zebrafish. *Dev. Biol.* **341**, 167–175 (2010).
24. Mueller, D. *et al.* Misguided transcriptional elongation causes mixed lineage leukemia. *PLoS Biol.* **7**, e1000249 (2009).
25. Bai, X. *et al.* TIF1 $\gamma$  controls erythroid cell fate by regulating transcription elongation. *Cell* **142**, 133–143 (2010).
26. Wheeler, G. N. & Brandli, A. W. Simple vertebrate models for chemical genetics and drug discovery screens: lessons from zebrafish and *Xenopus*. *Dev. Dyn.* **238**, 1287–1308 (2009).
27. Schiff, M. H., Strand, V., Oed, C. & Loew-Friedrich, I. Leflunomide: efficacy and safety in clinical trials for the treatment of rheumatoid arthritis. *Drugs Today (Barc.)* **36**, 383–394 (2000).
28. Flaherty, K. T. *et al.* Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.* **363**, 809–819 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank G. Bollag and P. Lin for supplying PLX4720. The melanoma xenografts were performed with the assistance of T. Venezia-Bowman. *In situ* hybridization probes for *crestin* were supplied by P. Henion. We thank S. Lacadie, C. Ceol, Y. Houvras, T. Bowman, X. Bai and R. Field for discussions and/or comments on the manuscript. This work was supported by the Howard Hughes Medical Institute and the National Cancer Institute (National Institutes of Health) (L.I.Z.), Aid for Cancer Research, the American Society for Clinical Oncology and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (National Institutes of Health) (R.M.W.). M.L.T. was a Biotechnology and Biological Sciences Research Council/Pfizer Industrial CASE award recipient.

**Author Contributions** R.M.W. and L.I.Z. planned the project. The chemical screen was performed by R.M.W., S.R., J.C., F.C., C.J.B., H.K.L. and S.D. The *Xenopus* work and initial identification of NSC210627 was performed by M.L.T. in the laboratory of G.N.W. The *mbp:mCherry* work was performed by R.M.W. and C.K. The human DHODH assay was performed by M.K. at Genzyme. The rat neural crest work was performed by J.M. in the laboratory of S.M. The ChIP-seq experiments and data analysis were performed by P.B.R. and C.Y.L. in the laboratory of R.A.Y. The ChIP-PCR assays were performed by P.B.R., R.M.W. and E.L. The zebrafish elongation and melanoma assays were performed by R.M.W. and J.C. Statistical analysis was performed by D.N. Human melanoma tissue microarray analysis was performed by S.G. All authors discussed the results and commented on the manuscript.

**Author Information** The microarray data discussed in this publication have been deposited in the NCBI Gene Expression Omnibus database under accession numbers GSE24526, GSE24527, GSE24528 and GSE24529. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details accompany the full-text HTML version of the paper at [www.nature.com/nature](http://www.nature.com/nature). Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.I.Z. ([zon@enders.tch.harvard.edu](mailto:zon@enders.tch.harvard.edu)).

# FAS and NF- $\kappa$ B signalling modulate dependence of lung cancers on mutant EGFR

Treva G. Bivona<sup>1</sup>, Haley Hieronymus<sup>1</sup>, Joel Parker<sup>2</sup>, Kenneth Chang<sup>3</sup>, Miquel Taron<sup>4,5</sup>, Rafael Rosell<sup>4,5</sup>, Philicia Moonsamy<sup>1</sup>, Kimberly Dahlman<sup>1</sup>, Vincent A. Miller<sup>6</sup>, Carlota Costa<sup>4,5</sup>, Gregory Hannon<sup>3,7</sup> & Charles L. Sawyers<sup>1,7</sup>

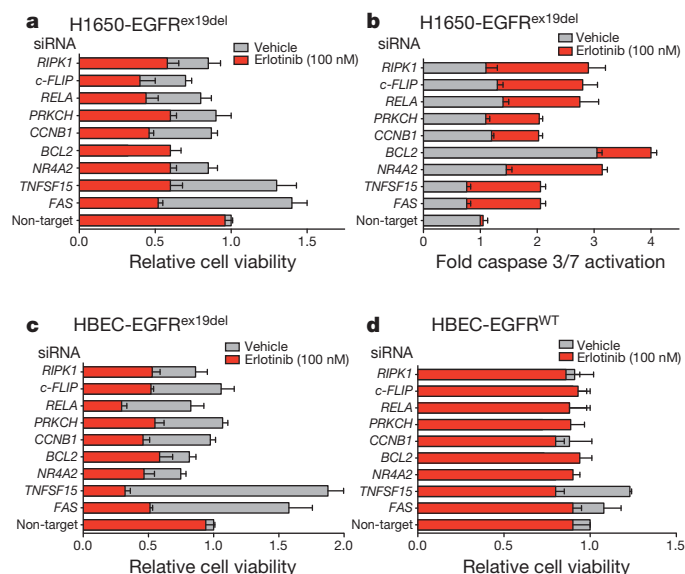
Human lung adenocarcinomas with activating mutations in EGFR (epidermal growth factor receptor) often respond to treatment with EGFR tyrosine kinase inhibitors (TKIs), but the magnitude of tumour regression is variable and transient<sup>1,2</sup>. This heterogeneity in treatment response could result from genetic modifiers that regulate the degree to which tumour cells are dependent on mutant EGFR. Through a pooled RNA interference screen, we show that knockdown of FAS and several components of the NF- $\kappa$ B pathway specifically enhanced cell death induced by the EGFR TKI erlotinib in EGFR-mutant lung cancer cells. Activation of NF- $\kappa$ B through overexpression of c-FLIP or IKK (also known as CFLAR and IKBKB, respectively), or silencing of I $\kappa$ B (also known as NFKBIA), rescued EGFR-mutant lung cancer cells from EGFR TKI treatment. Genetic or pharmacologic inhibition of NF- $\kappa$ B enhanced erlotinib-induced apoptosis in erlotinib-sensitive and erlotinib-resistant EGFR-mutant lung cancer models. Increased expression of the NF- $\kappa$ B inhibitor I $\kappa$ B predicted for improved response and survival in EGFR-mutant lung cancer patients treated with EGFR TKI. These data identify NF- $\kappa$ B as a potential companion drug target, together with EGFR, in EGFR-mutant lung cancers and provide insight into the mechanisms by which tumour cells escape from oncogene dependence.

Despite marked clinical successes achieved with inhibitors of 'driver' kinases that promote tumour growth, responses are rarely complete and also vary in duration<sup>3,4</sup>. We hypothesized that the heterogeneity of treatment response may result from genetic modifiers that regulate the degree to which tumour cells are dependent upon the driver kinase and the response to TKI treatment.

Using lung cancer as a model to identify such modifiers, we conducted a screen for genes that, when silenced, enhance EGFR dependence in EGFR-mutant lung cancer cell lines. We intentionally selected H1650 cells for the primary screen because they are insensitive to EGFR TKIs despite expressing a mutant EGFR with an exon 19 deletion (ex19del) that predicts for erlotinib sensitivity in patients. H1650 cells harbour no known EGFR TKI resistance mechanisms other than functional PTEN loss, which does not fully account for their insensitivity<sup>5</sup>. To identify small hairpin RNAs (shRNAs) that might restore dependence on mutant EGFR in H1650 cells, we introduced a pooled, shRNA library targeting >2,000 'cancer-relevant' genes<sup>6</sup> (Supplementary Table 1) and treated these cells with vehicle or erlotinib, a standard EGFR TKI used in EGFR-mutant lung cancer patients. Hairpins targeting 36 genes reproducibly conferred erlotinib sensitivity in H1650 cells across three replicates (threshold 1.5-fold depletion in erlotinib/vehicle,  $P < 0.1$ ; Supplementary Table 2).

Among the primary screen hits, 18 of the targeted genes, including FAS, could be directly or indirectly linked to NF- $\kappa$ B signalling, which is known to have a role in survival signalling (Supplementary Table 2). Because recent data demonstrated that FAS and NF- $\kappa$ B signalling can promote tumour growth<sup>7-9</sup>, we proposed that FAS-NF- $\kappa$ B may rescue

EGFR-mutant tumour cells from EGFR inhibition. First, we confirmed that shRNAs targeting six of the highest scoring NF- $\kappa$ B pathway-associated genes identified in the primary pooled screen effectively silenced expression of their targets in H1650 cells (Supplementary Fig. 1a). Then we validated their growth inhibitory effects in erlotinib-treated H1650 cells using independent siRNAs (Fig. 1a, Supplementary Fig. 1b). Importantly, this reduction in cell viability was associated with increased caspase 3/7 activity (Fig. 1b), indicating that knockdown of these genes promoted erlotinib-induced apoptosis. To address directly the role of the NF- $\kappa$ B pathway in EGFR TKI sensitivity, we knocked down the major NF- $\kappa$ B subunit RELA (not represented in the pooled library) and found that RELA knockdown also induced erlotinib sensitivity in H1650 cells using short term viability and apoptosis readouts as well as clonogenic assays (Fig. 1a, b and Supplementary Fig. 2). Furthermore, this sensitizing effect was specific to EGFR inhibition because stable RELA knockdown did not alter sensitivity to cisplatin, paclitaxel or ultraviolet treatment (Supplementary Fig. 2) or other TKI (imatinib) (Supplementary Fig. 3). Because c-FLIP and



**Figure 1 | Mutant EGFR oncogene dependence requires downregulation of the FAS-NF- $\kappa$ B pathway.** **a**, Viability (CellTiter-Glo assay) of H1650 cells treated with vehicle or 100 nM erlotinib upon introduction of either a non-target siRNA pool or gene-specific siRNA pools targeting the genes. Relative cell viability is fold change in viability in erlotinib relative to vehicle non-target siRNA control (viability decrease >25% = validated).  $n = 3$ ; mean + s.e.m. **b**, Caspase 3/7 activation (Caspase-Glo assay) in indicated cell lines treated as in **a**. **c**, **d**, Viability (CellTiter-Glo assay) of indicated isogenic HBEC cells treated as in **a**.  $n = 3$ ; mean + s.e.m.

<sup>1</sup>Human Oncology and Pathogenesis Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 20, New York, New York 10065, USA. <sup>2</sup>Expression Analysis, Inc., 4324 South Alston, Avenue, Durham, North Carolina 27713, USA. <sup>3</sup>Cold Spring Harbor Laboratory, One Bungtown Road, Cold Spring Harbor, New York 11724, USA. <sup>4</sup>Catalan Institute of Oncology, Hospital Germans Trias i Pujol, Carretera Del Canyet s/n, 08916 Barcelona, Spain. <sup>5</sup>Pangaea Biotech, USP Dexeus University Institute, Sabino Arana 5 - 19, 08028 Barcelona, Spain. <sup>6</sup>Thoracic Medical Oncology, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. <sup>7</sup>Howard Hughes Medical Institute, Chevy Chase, Maryland 20815, USA.



RIPK have been implicated as intermediate signalling proteins linking FAS to NF- $\kappa$ B<sup>10</sup>, we asked if these genes also regulate erlotinib sensitivity. Indeed, silencing of *c-FLIP* or RIPK induced erlotinib sensitivity in H1650 cells (Fig. 1a, b). The effects of the entire panel of siRNAs targeting nine different NF- $\kappa$ B pathway-associated genes were observed regardless of PTEN status (Supplementary Fig. 4a, b).

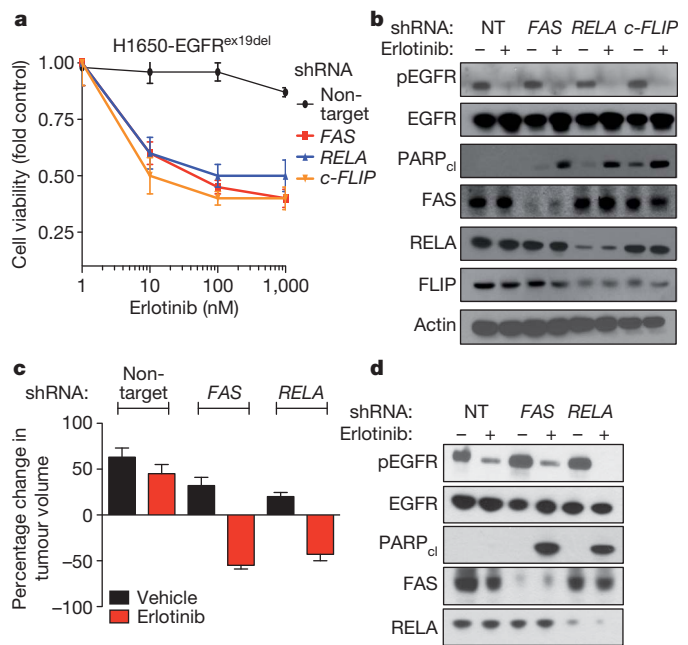
Next we asked whether the NF- $\kappa$ B pathway genes that scored in H1650 cells show similar effects in other EGFR mutant lung cancer cell lines. Whereas H1650 cells express EGFR<sup>ex19del</sup>, 11–18 human lung cancer cells express EGFR(L858R) yet are also relatively insensitive to erlotinib without a resistance mechanism validated in patients<sup>11</sup>. Each of the siRNAs that scored in H1650 cells also scored in cell viability and survival assays in 11–18 cells (Supplementary Fig. 4c, d). Silencing of the same set of genes also enhanced the growth suppressive effects of erlotinib in HCC827 cells (expressing EGFR<sup>ex19del</sup>) and in H3255 cells (expressing EGFR(L858R)), both of which are relatively more sensitive to EGFR TKIs (Supplementary Fig. 4e–h).

Because unknown alterations in human lung cancer cells could also influence erlotinib sensitivity, we used an isogenic system of EGFR-transduced human bronchial epithelial cells (HBEC) to test whether silencing of these genes cooperates with mutant EGFR to induce oncogene dependence. Consistent with prior data<sup>12</sup>, human bronchial epithelial cells (HBEC)-EGFR(L858R) and HBEC-EGFR<sup>ex19del</sup> cells were not sensitive to erlotinib (100 nM). However, silencing of each of the nine genes studied in the lung cancer cell lines also induced erlotinib sensitivity in HBEC-EGFR<sup>ex19del</sup> (Fig. 1c) and HBEC-EGFR(L858R) cells (Supplementary Fig. 5). Induction of erlotinib sensitivity seemed equivalent across both the exon19del and L858R EGFR genotypes. The erlotinib-sensitizing effect of silencing these genes was specific to mutant EGFR because no potentiating effect was seen in wild-type HBEC-EGFR<sup>WT</sup> cells (Fig. 1d).

Because FAS knockdown promoted erlotinib-induced apoptosis, we measured the activation state of three signalling pathways linked to cell survival (AKT, ERK, also known as MAPK1, and NF- $\kappa$ B) to determine which, if any, was associated with erlotinib-induced cell death. In erlotinib-sensitive HCC827 cells, erlotinib treatment alone led to reduced levels of phosphorylated pAKT, pERK and pRELA (a measure of NF- $\kappa$ B activity), regardless of FAS expression (Supplementary Fig. 6). But in erlotinib-resistant H1650 cells, these effects of erlotinib were only observed when FAS was silenced (Supplementary Fig. 6). In HBEC-EGFR(L858R) cells, erlotinib suppressed AKT and ERK activation regardless of FAS expression, but RELA phosphorylation was abolished only upon FAS silencing (Supplementary Fig. 6). Together these data implicate persistent NF- $\kappa$ B signalling in resistance to erlotinib-induced cell death.

NF- $\kappa$ B signalling was recently shown to be essential for KRAS-driven tumour growth<sup>13</sup>. Oncogenic KRAS and EGFR may drive tumour growth through a common signalling pathway. Because in our studies NF- $\kappa$ B knockdown alone was not lethal in EGFR-mutant lung cancer cell lines (HCC827, H3255), we compared the effect of silencing NF- $\kappa$ B in HBEC-KRAS12V cells versus HBEC-EGFR(L858R) cells directly. Knockdown of RELA alone impaired the growth of HBEC-KRAS12V but not HBEC-EGFR(L858R) cells (Supplementary Fig. 7), indicating that mutant EGFR does not function identically to mutant KRAS in this model.

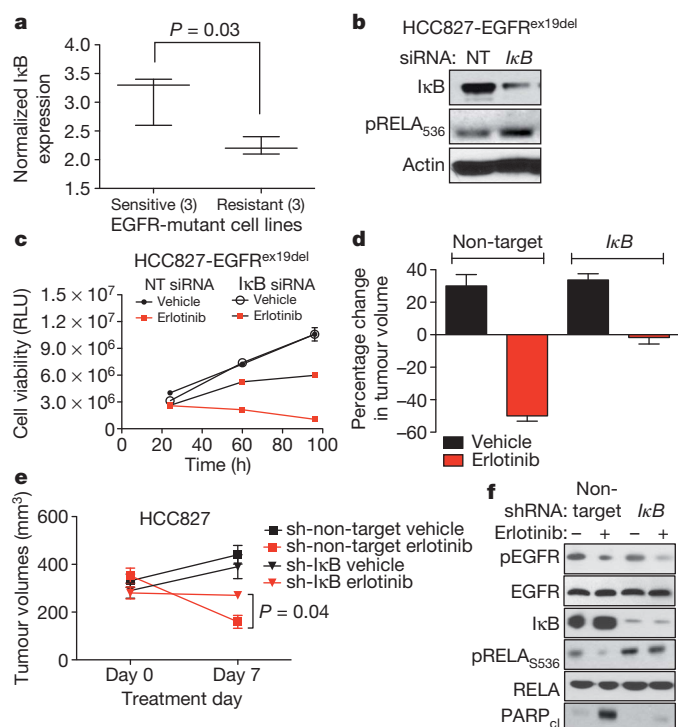
To extend our findings to *in vivo* models, we used shRNAs targeting FAS, RELA or *c-FLIP* (distinct from the primary screen) to achieve stable knockdown (Supplementary Fig. 8) of the cognate protein in erlotinib-resistant H1650 cells. Silencing of each gene by stable shRNA enhanced erlotinib sensitivity, with a 1–2 log-fold decrease in IC<sub>50</sub> (half maximal inhibitory concentration; Fig. 2a) and induction of apoptosis as measured by PARP cleavage (Fig. 2b). Similar results were observed in 11–18 and HBEC-EGFR(L858R) cells (Supplementary Fig. 9). We found that erlotinib treatment induced tumour regression and apoptosis only upon knockdown of FAS or RELA (Fig. 2c, d and Supplementary Fig. 10a, b) in H1650 human xenograft tumours.



**Figure 2 | Suppression of the FAS-NF- $\kappa$ B pathway enhances EGFR TKI response in EGFR-mutant cells and tumour models. a**, Erlotinib dose response in H1650 cells expressing a non-target control shRNA or a FAS (shRNA2), RELA (shRNA1), or *c-FLIP* (shRNA1) shRNA (5  $\mu$ M erlotinib IC<sub>50</sub> in parental H1650 cells). Cell viability was assayed as in Fig. 1. *n* = 3; mean  $\pm$  s.e.m. **b**, Immunoblots showing expression of indicated signalling proteins in H1650 cells generated in **a**. Data represent three independent experiments. NT, non-target; PARP<sub>cl</sub>, PARP cleaved. (The decrease in *c-FLIP* protein by shRNA was less complete than the decrease in *c-FLIP* mRNA level by siRNA; Supplementary Fig. 1). **c**, Effects of stable knockdown of FAS (shRNA3) or RELA (shRNA2) on erlotinib sensitivity in H1650 xenograft tumours, compared to non-target shRNA control H1650 tumours. Established tumours (>200 mm<sup>3</sup>, *n* = 10 per treatment group) were randomized and treated for 7 days with 12.5 mg erlotinib per kg per day, *n* = 10. Data expressed as mean  $\pm$  s.e.m. **d**, Immunoblots showing expression of indicated proteins in representative H1650 tumour xenografts from **c** analysed at treatment day 7.

We next asked whether activation of NF- $\kappa$ B might be sufficient to confer erlotinib resistance in EGFR-mutant tumours. Canonical NF- $\kappa$ B signalling requires downregulation of the NF- $\kappa$ B inhibitor I $\kappa$ B<sup>14</sup>. Therefore we predicted that decreased I $\kappa$ B levels, leading to increased NF- $\kappa$ B signalling, might promote EGFR TKI resistance. First, we noted that I $\kappa$ B expression was lower in three erlotinib-resistant EGFR-mutant lung cancer cell lines that are resistant to erlotinib compared with three sensitive cell lines (Fig. 3a and Supplementary Fig. 11). Furthermore, we found that knockdown of I $\kappa$ B increased NF- $\kappa$ B phosphorylation (Fig. 3b) and conferred partial resistance to erlotinib in HCC827 cells *in vitro* (Fig. 3c) and *in vivo* (Fig. 3d–f). These effects were specific to EGFR inhibition because stable knockdown of I $\kappa$ B did not protect HCC827 cells from cisplatin, paclitaxel or ultraviolet treatment, as measured by clonogenic assays (Supplementary Fig. 12). *c-FLIP* has been implicated as an intermediate signalling molecule that activates NF- $\kappa$ B downstream of FAS<sup>10</sup>. Ectopic expression of *c-FLIP* induced persistent phosphorylation of RELA and resistance to erlotinib in HCC827 cells (Supplementary Fig. 13). Together these data suggest that NF- $\kappa$ B activation can rescue EGFR-mutant tumour cells from EGFR TKI.

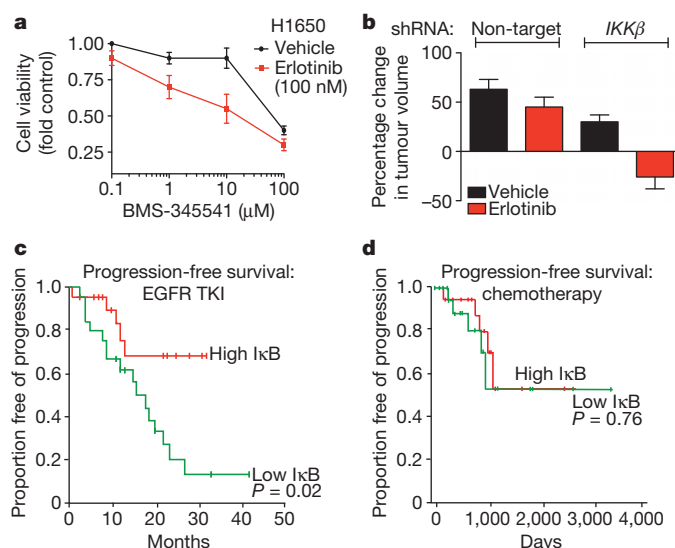
Acquired resistance to erlotinib in patients can occur through second site, drug-resistant mutations in EGFR or through amplification of the MET kinase<sup>15,16</sup>. These same resistance mechanisms can also evolve *in vitro* with prolonged growth in EGFR TKI<sup>17,18</sup>. We therefore asked if NF- $\kappa$ B activation occurs naturally during the derivation of erlotinib-resistant HCC827 cells through prolonged and continuous EGFR TKI treatment. In parental HCC827 cells acute erlotinib treatment (100 nM,



**Figure 3 | NF- $\kappa$ B activation through I $\kappa$ B downregulation confers EGFR TKI resistance in EGFR-mutant lung cancer models.** **a**, Correlation of I $\kappa$ B expression with EGFR TKI sensitivity in EGFR-mutant lung cancer cells (I $\kappa$ B mRNA expression from OncoPrint; sensitive  $<0.02 \mu\text{M}$  IC<sub>50</sub>: HCC827, H3255, HCC4006; resistant  $>1 \mu\text{M}$  IC<sub>50</sub>: H1650, H1975, H820). **b**, Immunoblots showing I $\kappa$ B and pRELA expression in lysates from HCC827 treated with non-targeting or I $\kappa$ B siRNA pools. **c**, Viability (CellTiter-Glo assay) of HCC827 cells treated with non-target siRNA pool or I $\kappa$ B siRNA pool and either vehicle or erlotinib (100 nM). RLU is relative luciferase units ( $n = 3$ ; mean  $\pm$  s.e.m.). **d**, **e**, Effects of stable knockdown of I $\kappa$ B on erlotinib sensitivity in HCC827 tumour xenografts compared to non-target shRNA control HCC827 tumours. Tumours were established and treated as in Fig. 2c ( $n = 8$  per treatment group, mean  $\pm$  s.e.m., 7-day treatment). **f**, Immunoblots showing expression of indicated proteins in representative HCC827 tumour xenografts analysed at treatment day 7.

24 h) inhibited EGFR, MET, AKT, ERK and RELA phosphorylation. Continuous erlotinib treatment resulted in the outgrowth of eight resistant HCC827 subclones (over  $>8$  weeks) in which EGFR phosphorylation remained inhibited by erlotinib. Consistent with prior data<sup>18</sup>, we observed compensatory upregulation of phosphorylated MET in four erlotinib-resistant subclones. We also observed increased FAS expression in the absence of MET upregulation in four additional subclones. Increased RELA phosphorylation and decreased I $\kappa$ B expression were observed in three out of four sub-clones in which FAS was upregulated (Supplementary Fig. 14a). Knockdown of FAS in those subclones with FAS upregulation decreased pRELA and enhanced erlotinib sensitivity, but did not result in decreased pAKT or pERK (Supplementary Fig. 14b, c). Thus, FAS-mediated EGFR TKI resistance is distinct from MET-mediated EGFR TKI resistance, which occurs primarily through reactivation of pERK and pAKT in HCC827 cells<sup>18</sup>. Our results are consistent with recent data establishing both MEK-dependent<sup>19</sup> and MEK-independent<sup>20</sup> mechanisms of resistance of BRAF-mutant melanomas to the BRAF inhibitor PLX4032.

NF- $\kappa$ B pathway inhibition can be achieved through inhibition of I $\kappa$ B kinase (IKK $\beta$ , also known as I $\kappa$ BKB), the primary kinase that promotes I $\kappa$ B instability leading to NF- $\kappa$ B activation<sup>14</sup>. We therefore asked if pharmacologic or genetic inhibition of IKK $\beta$  enhances erlotinib sensitivity in our erlotinib-resistant lung cancer models. Treatment of H1650 (Fig. 4a) or HBEC-EGFR(L858R) cells (Supplementary Fig. 15a, b) with the IKK $\beta$  inhibitor BMS-345541 (ref. 21) inhibited



**Figure 4 | Rationale for combined NF- $\kappa$ B and EGFR inhibition in EGFR-mutant lung cancers.** **a**, Dose response in H1650 cells treated with BMS-345541 (IKK inhibitor) and additionally either vehicle or erlotinib (100 nM). Viability was measured as in Fig. 1 ( $n = 3$ , mean  $\pm$  s.e.m.). **b**, Effects of stable knockdown of IKK $\beta$  on erlotinib sensitivity in H1650 tumour xenografts compared to non-target shRNA control H1650 tumours. Established tumours ( $>200 \text{ mm}^3$ ,  $n = 10$  per treatment group) were randomized and treated for 7 days with 12.5 mg erlotinib per kg per day or vehicle. Data are expressed as in Fig. 2c ( $\pm$  s.e.m.). **c**, **d**, Effects of I $\kappa$ B expression on progression free survival in patients with EGFR-mutant lung cancers (**c**) treated with single agent EGFR TKI ( $n = 52$ ) or (**d**) chemotherapy and surgery ( $n = 43$ ). Clinical characteristics and responses were defined previously<sup>22</sup>. Median progression-free survival and overall survival for the entire EGFR TKI-treated cohort were 20 months (95% confidence interval, 13–26.9) and 33 months (95% confidence interval, 22.2–43.8), respectively.

RELA phosphorylation and restored erlotinib sensitivity. Because BMS-345541 has suboptimal pharmacokinetic properties we used an shRNA to knock down IKK $\beta$  in H1650 cells. The resulting H1650 cells, or those transduced with a non-target control shRNA, were then grown as xenograft tumours and treated with either vehicle or erlotinib. IKK $\beta$  knockdown resulted in increased levels of I $\kappa$ B and decreased NF- $\kappa$ B phosphorylation, as expected. Erlotinib treatment induced tumour regression and apoptosis only upon knockdown of IKK $\beta$  in this system (Fig. 4b and Supplementary Fig. 15c, d).

To determine the clinical relevance of our findings, we examined the status of NF- $\kappa$ B activation in a cohort of 52 patients (part of a previously characterized cohort<sup>22</sup>) with EGFR mutant lung cancer who were treated with erlotinib and did not have evidence of the T790M mutation as a potential cause of erlotinib resistance. Because reduced I $\kappa$ B levels were associated with erlotinib resistance in cell lines and tumour models, we asked if I $\kappa$ B expression correlated with EGFR TKI response in patients with EGFR-mutant lung cancers. Low I $\kappa$ B expression ('high-NF- $\kappa$ B' activation state) was predictive of worse progression-free survival (Fig. 4c and Supplementary Table 4a) and decreased overall survival (Supplementary Fig. 16a and Supplementary Table 4b). This finding was specific to erlotinib because I $\kappa$ B expression did not predict outcome in a previously characterized cohort<sup>23</sup> of 43 EGFR-mutant lung cancer patients that were not treated with EGFR TKIs but instead with chemotherapy and surgery (Fig. 4d and Supplementary Fig. 16b). We also observed a significant positive correlation between FAS expression and RELA, c-FLIP and I $\kappa$ B levels in the erlotinib-treated patients (Supplementary Table 5). Together, our findings suggest that the extent to which EGFR-mutant tumour cells engage the NF- $\kappa$ B pathway may, in part, explain the non-uniform response to EGFR TKI treatment observed in EGFR-mutant lung cancer patients and provide rationale for testing an IKK inhibitor in combination with an EGFR TKI in EGFR-mutant lung cancer patients.

## METHODS SUMMARY

**Compounds and cell culture.** H1650, H11-18, H3255, HCC827 cells were provided by W. Pao and were grown in RPMI supplemented with 10% FBS. H1650 PTEN-reconstituted cells have been described previously<sup>5</sup>. HBEc cell lines were provided by J. Minna and grown in K-SFM. Erlotinib was from LC Labs. BMS-354541 was from Sigma-Aldrich.

**Antibodies and immunoblots.** All antibodies were from Cell Signaling Technologies and used at 1:1,000 dilution according to manufacturer's instructions. Lysates from cell lines and xenograft tumours were generated using standard methods and RIPA buffer and assayed by immunoblots.

**Pooled shRNA screen.** H1650 cells were infected with 6,783 shRNAs targeting ~2,500 cancer-associated genes in a pooled format. Abundance of each hairpin within the total population over time was detected by hybridization to custom Agilent oligonucleotide microarrays with half-hairpin and barcode probes corresponding to each shRNA<sup>6</sup>. The abundance difference for the erlotinib-synthetic screen was determined as the average of the log<sub>2</sub> signal from erlotinib treated cells at later time points (day 10 and day 21) minus the average of this log<sub>2</sub> signal from vehicle treated cells at the later time points (day 10 and day 21). Significant hits were defined as those altered by 1.5 fold,  $P < 0.1$ .

**RNA interference assays.** Sequences for individual small interfering RNA (siRNA) or shRNA (from Sigma or Open Biosystems) are shown in Supplementary Table 3. For siRNA experiments, cells were treated in triplicate according to manufacturer's protocol. For shRNA experiments, cells were spin-infected with shRNA virus. Where indicated cells were treated with vehicle or erlotinib 24 h after RNAi treatment. Gene silencing was confirmed by mRNA isolation and gene-specific quantitative RT-PCR and by immunoblots. Cell viability or survival was measured using CellTiter-Glo and Caspase-Glo, respectively, 24–96 h after drug treatment according to the manufacturer's instructions and normalized to cells with a non-target siRNA pool or shRNA hairpin.

Received 2 June 2010; accepted 24 January 2011.

- Haber, D. A. *et al.* Molecular targeted therapy of lung cancer: EGFR mutations and response to EGFR inhibitors. *Cold Spring Harb. Symp. Quant. Biol.* **70**, 419–426 (2005).
- Miller, V. A. *et al.* Molecular characteristics of bronchioloalveolar carcinoma and adenocarcinoma, bronchioloalveolar carcinoma subtype, predict response to erlotinib. *J. Clin. Oncol.* **26**, 1472–1478 (2008).
- Weinstein, I. B. & Joe, A. Oncogene addiction. *Cancer Res.* **68**, 3077–3080 (2008).
- Knight, Z. A., Lin, H. & Shokat, K. M. Targeting the cancer kinome through polypharmacology. *Nature Rev. Cancer* **10**, 130–137 (2010).
- Sos, M. L. *et al.* PTEN loss contributes to erlotinib resistance in EGFR-mutant lung cancer by activation of Akt and EGFR. *Cancer Res.* **69**, 3256–3261 (2009).
- Silva, J. M. *et al.* Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science* **319**, 617–620 (2008).
- Peter, M. E. *et al.* The CD95 receptor: apoptosis revisited. *Cell* **129**, 447–450 (2007).
- Chen, L. *et al.* CD95 promotes tumour growth. *Nature* **465**, 492–496 (2010).
- O'Reilly, L. A. *et al.* Membrane-bound Fas ligand only is essential for Fas-induced apoptosis. *Nature* **461**, 659–663 (2009).
- Green, D. R. Cancer: a wolf in wolf's clothing. *Nature* **465**, 433 (2010).
- Gong, Y. *et al.* High expression levels of total IGF-1R and sensitivity of NSCLC cells *in vitro* to an anti-IGF-1R antibody (R1507). *PLoS ONE* **4**, e7273 (2009).
- Sato, M. *et al.* Multiple oncogenic changes (K-RAS<sup>V12</sup>, p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res.* **66**, 2116–2128 (2006).
- Meylan, E. *et al.* Requirement for NF- $\kappa$ B signalling in a mouse model of lung adenocarcinoma. *Nature* **462**, 104–107 (2009).
- Luo, J. L., Kamata, H. & Karin, M. IKK/NF- $\kappa$ B signaling: balancing life and death—a new approach to cancer therapy. *J. Clin. Invest.* **115**, 2625–2632 (2005).
- Pao, W. *et al.* Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).
- Bean, J. *et al.* MET amplification occurs with or without T790M mutations in EGFR mutant lung tumors with acquired resistance to gefitinib or erlotinib. *Proc. Natl Acad. Sci. USA* **104**, 20932–20937 (2007).
- Engelman, J. A. *et al.* Allelic dilution obscures detection of a biologically significant resistance mutation in EGFR-amplified lung cancer. *J. Clin. Invest.* **116**, 2695–2706 (2006).
- Turke, A. B. *et al.* Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC. *Cancer Cell* **17**, 77–88 (2010).
- Johannessen, C. M. *et al.* COT drives resistance to RAF inhibition through MAP kinase pathway reactivation. *Nature* **468**, 968–972 (2010).
- Nazarian, R. *et al.* Melanomas acquire resistance to B-Raf(V600E) inhibition by RTK or N-RAS upregulation. *Nature* **468**, 973–977 (2010).
- Yang, J., Amiri, K. I., Burke, J. R., Schmid, J. A. & Richmond, A. BMS-345541 targets inhibitor of  $\kappa$ B kinase and induces apoptosis in melanoma: involvement of nuclear factor  $\kappa$ B and mitochondria pathways. *Clin. Cancer Res.* **12**, 950–960 (2006).
- Rosell, R. *et al.* Screening for epidermal growth factor receptor mutations in lung cancer. *N. Engl. J. Med.* **361**, 958–967 (2009).
- Chitale, D. *et al.* An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* **28**, 2773–2783 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank W. Pao, H. Varmus and members of the MSKCC Lung Cancer Oncogenome group for intellectual support. We thank J. Javier Sanchez for analysing human lung cancer gene expression data. We thank J. Wongvipat and E. Philips for technical assistance and W. Polkinghorn and members of the Sawyers Laboratory for critique of the manuscript. T.G.B. is supported by the MSKCC Clinical Scholars Fellowship funded by the Charles A. Dana Foundation, an ASCO YIA, and the Caine Halter Lung Cancer Research Fund/Uniting Against Lung Cancer Research Grant. C.L.S. and G.H. are Investigators of the Howard Hughes Medical Institute.

**Author Contributions** T.G.B. and H.H. designed research, performed experiments, analysed data and co-wrote the paper. K.C., P.M., K.D., M.T. and C.C. provided reagents, performed experiments and analysed data. J.P. and V.A.M. analysed data. R.R., G.H. and C.L.S. designed research, analysed experiments and co-wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to C.L.S. ([sawyers@mskcc.org](mailto:sawyers@mskcc.org)).



# A *cis*-regulatory map of the *Drosophila* genome

Nicolas Nègre<sup>1\*</sup>, Christopher D. Brown<sup>1\*</sup>, Lijia Ma<sup>1\*</sup>, Christopher Aaron Bristow<sup>2\*</sup>, Steven W. Miller<sup>3\*</sup>, Ulrich Wagner<sup>4\*</sup>, Pouya Kheradpour<sup>2</sup>, Matthew L. Eaton<sup>5</sup>, Paul Loriaux<sup>6</sup>, Rachel Sealfon<sup>2</sup>, Zirong Li<sup>4</sup>, Haruhiko Ishii<sup>3</sup>, Rebecca F. Spokony<sup>1</sup>, Jia Chen<sup>7</sup>, Lindsay Hwang<sup>4</sup>, Chao Cheng<sup>8,9,10</sup>, Richard P. Auburn<sup>11</sup>, Melissa B. Davis<sup>1</sup>, Marc Domanus<sup>1</sup>, Parantu K. Shah<sup>12</sup>, Carolyn A. Morrison<sup>1</sup>, Jennifer Zieba<sup>1</sup>, Sarah Suchy<sup>1</sup>, Lionel Senderowicz<sup>1</sup>, Alec Victorsen<sup>1</sup>, Nicholas A. Bild<sup>1</sup>, A. Jason Grundstad<sup>1</sup>, David Hanley<sup>7</sup>, David M. MacAlpine<sup>5</sup>, Mattias Mannervik<sup>13</sup>, Koen Venken<sup>14</sup>, Hugo Bellen<sup>14</sup>, Robert White<sup>15</sup>, Mark Gerstein<sup>8,9</sup>, Steven Russell<sup>11</sup>, Robert L. Grossman<sup>1,7,16</sup>, Bing Ren<sup>4,17</sup>, James W. Posakony<sup>3</sup>, Manolis Kellis<sup>2</sup> & Kevin P. White<sup>1</sup>

**Systematic annotation of gene regulatory elements is a major challenge in genome science. Direct mapping of chromatin modification marks and transcriptional factor binding sites genome-wide<sup>1,2</sup> has successfully identified specific subtypes of regulatory elements<sup>3</sup>. In *Drosophila* several pioneering studies have provided genome-wide identification of Polycomb response elements<sup>4</sup>, chromatin states<sup>5</sup>, transcription factor binding sites<sup>6–9</sup>, RNA polymerase II regulation<sup>8</sup> and insulator elements<sup>10</sup>; however, comprehensive annotation of the regulatory genome remains a significant challenge. Here we describe results from the modENCODE *cis*-regulatory annotation project. We produced a map of the *Drosophila melanogaster* regulatory genome on the basis of more than 300 chromatin immunoprecipitation data sets for eight chromatin features, five histone deacetylases and thirty-eight site-specific transcription factors at different stages of development. Using these data we inferred more than 20,000 candidate regulatory elements and validated a subset of predictions for promoters, enhancers and insulators *in vivo*. We identified also nearly 2,000 genomic regions of dense transcription factor binding associated with chromatin activity and accessibility. We discovered hundreds of new transcription factor co-binding relationships and defined a transcription factor network with over 800 potential regulatory relationships.**

To reveal chromatin, promoter and enhancer domains in the genome, we performed a developmental time course of six histone modifications, the *Drosophila* CREB binding protein (CBP) and RNA polymerase II (PolII) across twelve stages of embryonic, larval, pupal and adult development (Supplementary Table 1 and Supplementary Figs 1–2; see Supplementary Methods). We used whole animals to generate the maximum number of chromatin marks across the genome. We identified 506,001 chromatin-associated features covering 101 megabases (Mb) (86.99%) of the non-repetitive genome. To relate these chromatin features to gene activity, we quantified transcript levels by high-throughput complementary DNA sequencing (RNA-seq) with the same biological samples used for chromatin immunoprecipitation (ChIP). Additionally, we mapped 38 functionally diverse transcription factors in different developmental stages and cell types. A total of 155,048 transcription factor binding sites (TFBSs) were identified, including 35,096 unique TFBSs. Of these, 93.76% overlap at least one chromatin feature. We noted that although the majority of factors are bound in discrete regions, several are distributed in larger domains

(Supplementary Table 1 and Supplementary Fig. 3). We also characterized the binding distributions of five Histone deacetylases (HDACs), identifying a total of 19,937 HDAC binding sites mapping to 7,692 unique genomic locations. Of these, 99.25% overlap with at least one chromatin feature, and 94.58% overlap with at least one TFBS. All data from this study are available at <http://www.modencode.org> and <http://www.cistrack.org>.

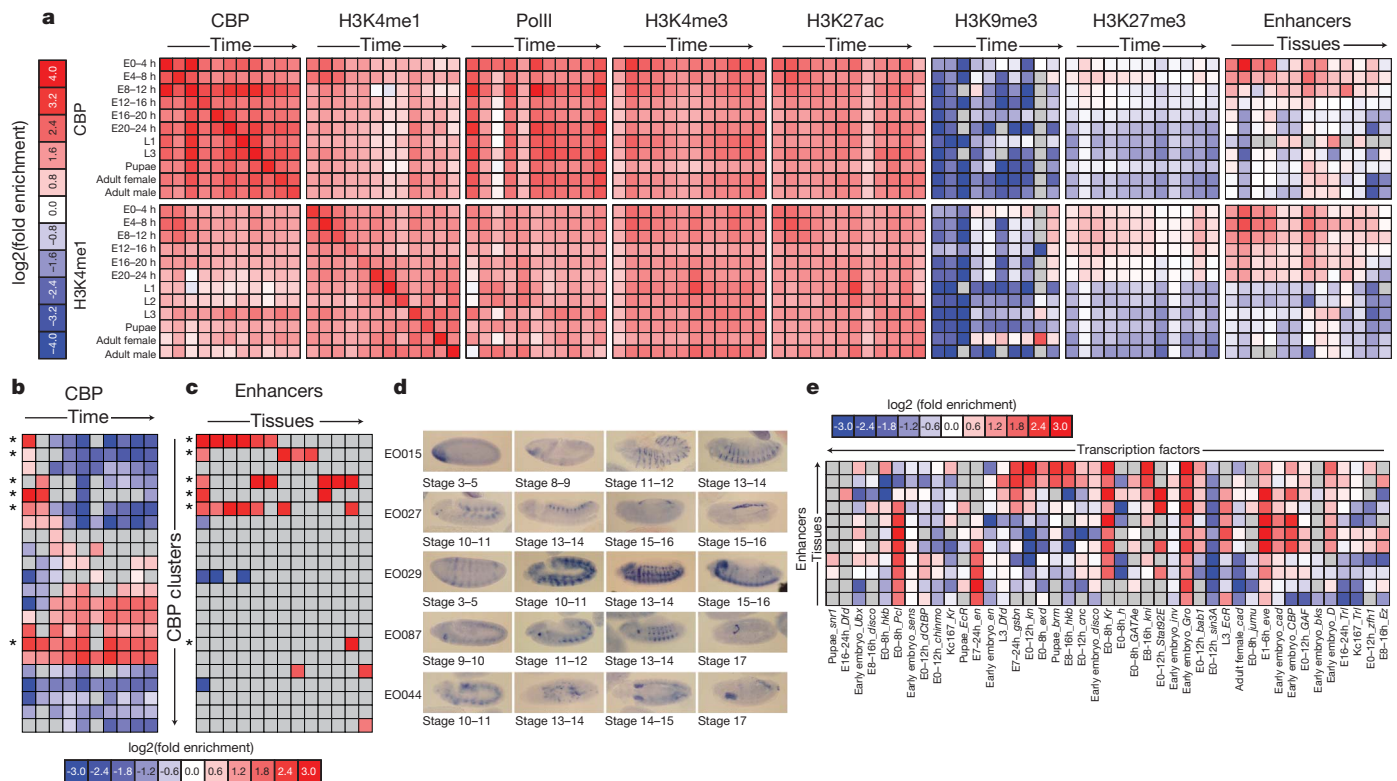
For each chromatin mark, very few target genes showed either repressive or activating marks across all of development; most genes were within dynamically marked regions (Supplementary Fig. 4a). We observed three major patterns of chromatin mark distributions, corresponding to active promoters (H3K4me3, H3K9ac, H3K27ac), repressive states and silencers (H3K27me3, PHO/Polycomb response elements (PREs)), and enhancers (CBP, H3K4me1) (Supplementary Fig. 4b).

The first pattern, represented by the activating histone modifications H3K4me3, H3K9ac and H3K27ac, was strongly associated with transcription start sites (TSSs) and was positively correlated with gene expression levels<sup>11</sup> (Supplementary Figs 4b, 5a, 7). Although the enrichment of activating histone modifications was quite marked, we note that a substantial fraction of genes (34%) were expressed but lacked H3K4me3 marks at their annotated TSS (Supplementary Figs 8–12). Regions marked by each activating modification also significantly overlapped with class I insulators, PolII binding sites, and a large fraction of TFBSs (Supplementary Fig. 1).

The second type of pattern, repressive chromatin marks H3K9me3 and H3K27me3, showed a distribution of large domains throughout development (Supplementary Figs 2, 6, 13). As expected on the basis of polytene chromosomes *in situ* data, H3K9me3 marks localized to ~20 developmentally stable domains primarily at centromeres<sup>12</sup>. H3K27me3 marks, in contrast, were remarkably dynamic (Fig. 1a). Dynamic domains may be due to changes in specific cell populations during development or the active addition and removal of H3K27me3 marks. Previous studies have implicated H3K27me3 dynamics in the regulation of homeotic genes, in the differentiation of stem cells, and in developmental processes in vertebrates<sup>13</sup>. We found between 123 and 438 discrete domains present at the developmental stages assayed, each with an average length of ~70 kb (Supplementary Fig. 13a and Supplementary Table 2). One-thousand two-hundred and sixty-four genes were associated with H3K27me3 in at least one stage of development, with 397

<sup>1</sup>Institute for Genomics and Systems Biology, Department of Human Genetics, The University of Chicago, 900 East 57th Street, Chicago, Illinois 60637, USA. <sup>2</sup>Massachusetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Division of Biological Sciences/CDB, University of California San Diego, La Jolla, California 92093, USA. <sup>4</sup>Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, California 92093-0653, USA. <sup>5</sup>Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>6</sup>Signaling Systems Laboratory, Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. <sup>7</sup>National Center for Data Mining, University of Illinois at Chicago, 851 S. Morgan Street, Chicago, Illinois 60607, USA. <sup>8</sup>Program in Computational Biology & Bioinformatics, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. <sup>9</sup>Department of Molecular Biophysics & Biochemistry, Yale University, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. <sup>10</sup>Department of Computer Science, Yale University, Bass 432, 266 Whitney Avenue, New Haven, Connecticut 06520, USA. <sup>11</sup>Department of Genetics and Cambridge Systems Biology Centre, University of Cambridge, Cambridge CB2 3EH, UK. <sup>12</sup>Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, Massachusetts 02115, USA. <sup>13</sup>Department of Developmental Biology, Wenner-Gren Institute, Arrhenius Laboratories E3, Stockholm University, S-106 91 Stockholm, Sweden. <sup>14</sup>Department of Molecular and Human Genetics, BCM, Houston, Texas 77030, USA. <sup>15</sup>Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge CB2 3DY, UK. <sup>16</sup>Department of Medicine, University of Chicago, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA. <sup>17</sup>Department of Cellular and Molecular Medicine, Institute of Genomic Medicine & Moores Cancer Center, 9500 Gilman Drive, La Jolla, California 92093, USA.

\*These authors contributed equally to this work.



**Figure 1 | Chromatin dynamics across *Drosophila* development.**

**a**, Enrichment of CBP and H3K4me1 (rows) within regions marked by other chromatin modifications, factors, or annotated enhancers (columns). Note that (1) CBP is enriched within all active marks (H3K4me3, H3K27ac, H3K9ac, H3K4me1 and PolII) at all stages of development, and (2) early embryo (0–16 h) CBP- and H3K4me1-marked regions are enriched within H3K27me3 domains and annotated enhancers (right panel). **b**, Heatmap depicting fold enrichment of CBP-bound regions (columns) at different developmental stages for each of the 22 clusters of TSS-distal regions (rows) grouped by their protein binding profiles. A subset of the clusters shows significant enrichment for CBP at different developmental stages. **c**, Enrichment of enhancer categories

genes (31%) in domains present in all stages of development and 867 genes (69%) in dynamic domains (Fig. 1a). Stable H3K27me3 domains corresponded to those reported in embryos and tissue culture cells<sup>4</sup>, and were enriched for genes involved in development, transcription and segmentation. However, identification of stage-specific H3K27me3 domains revealed previously unappreciated H3K27me3 targets, including genes that control apoptosis, regulation of growth and neurotransmitter transport (Supplementary Fig. 14). We found that stable H3K27me3 domains are highly enriched for genes that exhibit stage- and tissue-specific expression, and are depleted for ubiquitously expressed genes (Supplementary Fig. 15).

H3K4me1-marked and CBP/p300-bound regions form a third, intermediate class of genomic elements known to be associated with active enhancers<sup>3,14</sup> (Fig. 1a). They were also associated with active promoters, activating histone marks and transcription factor binding sites (Supplementary Fig. 1). H3K4me1 and CBP were bound broadly across TSSs, typically positioned 1–2 kb upstream and downstream of the TSS, consistent with previous observations<sup>11</sup> (Supplementary Figs 2, 4b). Accordingly, these patterns were very dynamic across development (Supplementary Fig. 4a).

To characterize the regulators of chromatin mark dynamics, we determined the genome-wide distribution pattern of all five known *Drosophila* HDACs (HDAC1 (also known as RPD3), HDAC3, HDAC4, HDAC6 and HDACX (also known as HDAC11)). All five HDACs are enriched at active promoters, and enrichment is correlated with target gene expression level (Supplementary Fig. 16). HDAC4 and HDAC1/RPD3 binding sites also mark PREs. HDAC1 and HDAC4

(columns) for each of the 22 clusters of TSS-distal regions (rows). Many clusters enriched for CBP binding in early development are also strongly enriched for enhancers (rows marked with an asterisk). **d**, Embryo-specific CBP binding predicts unannotated enhancers. RNA *in situ* hybridizations with a Gal4 probe were used to stain transgenic embryos representing five different enhancer predictions (rows), at four to five different stages (columns). EO044 overlaps the known expression pattern for the neighbouring gene, *CG8745* (FlyExpress Database). **e**, Enrichment of enhancer annotations (rows) within the binding sites of each transcription factor (columns). For panels **a**, **b**, **c** and **e** grey boxes indicate no overlap. For panels **a**, **b**, **c** and **e** all values greater or less than zero are significant, false discovery rate (FDR) < 0.01.

binding sites are frequently found within H3K27me3 repressive domains (Supplementary Figs 17, 18), and are significantly enriched at embryonic PHO (a PcG recruiter protein)-bound regions (Supplementary Fig. 16f). Of the 537 HDAC1 and 4a binding sites that overlap H3K27me3 but not H3K4me3 (Supplementary Fig. 16), 149 overlap with 350 previously described<sup>15</sup> embryonic PHO sites (Supplementary Figs 17, 18). HDAC3 is primarily associated with transcribed, H3K36me3 marked exons<sup>16</sup> (Supplementary Fig. 16a, d).

Using the dynamic chromatin signatures and RNA-seq data, we sought next to systematically annotate *cis*-regulatory elements. To identify novel promoters, we identified coincident H3K4me3, PolII and RNA signals at least 1,000 base pairs away from any annotated TSS (see Supplementary Methods). In each developmental stage we found several hundred such regions (average, 485; range, 179–885), resulting in a total of 2,307 novel promoter predictions; 1,117 of which are supported by modENCODE cap analysis of gene expression (CAGE) data from embryos<sup>17</sup> (Supplementary Fig. 5a). We subjected 110 novel promoter predictions to biological validation using a luciferase reporter assay in Kc167 cells. Seventy-five of these 110 predicted promoters (68%) yielded significant luciferase activity in at least one orientation, with 26 displaying bi-directionality (Supplementary Fig. 5b and Supplementary Table 3). Together, the CAGE and reporter assay data indicate that a high proportion of these novel promoter predictions indeed correspond to previously unannotated TSSs.

To identify additional putative *cis*-regulatory elements on a genome-wide scale, we examined two signatures of enhancers, H3K4me1 and CBP/p300<sup>11,14</sup>. CBP and H3K4me1 are significantly enriched within

several classes of known enhancers from the CRM activity database (CAD)<sup>9</sup>. For example, we found a 15-fold ( $z$ -score of 26) and 5.9-fold ( $z$ -score of 10) enrichment for CBP and H3K4me1 overlap, respectively, with blastoderm-specific enhancers, indicating that our dynamic chromatin map successfully recovers previously annotated enhancers (Supplementary Fig. 19). Given that CBP can be recruited to enhancers by bound transcription factors, we sought to further support the functional relevance of CBP-bound regions by examining clusters of co-occupancy with other transcription factors. Several CBP clusters are bound by transcription factors known to interact physically with CBP, such as Bicoid, Dorsal (DL) and Trithorax-like (TRL (also known as GAF)); whereas other clusters are enriched for known enhancers (Fig. 1e) and are strongly enriched in K3K4me1 and the repressive mark H3K27me3 (Supplementary Fig. 20). In total, 14,450 distinct putative CBP-bound *cis*-regulatory elements were identified across the genome (Supplementary Table 18).

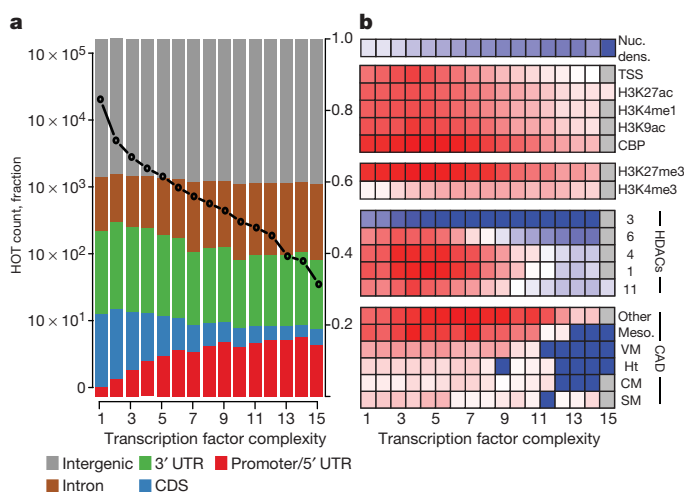
To validate the ability of CBP binding data to accurately identify *cis*-regulatory elements, we tested 33 putative enhancer sequences using reporter assays in transgenic *Drosophila*. We focused on putative enhancers that have dynamic CBP association during embryogenesis (Supplementary Table 4). Thirty of the 33 predicted enhancers produce specific reporter expression patterns (Fig. 1d and Supplementary Fig. 21). We also selected a set of putative insulator elements<sup>10</sup>, and we tested their activity in an enhancer-blocking assay based on the *eve* stripe 2 and 3 enhancers. We assayed a set of 15 genomic fragments associated with the binding of Centrosomal protein 190 kDa (CP190) + CTCF (class I), CP190 + suppressor of Hairy wing (Su(Hw)) (class II) and TRL<sup>10</sup>. We found that five of eight CTCF sites showed strong enhancer-blocking activity and the remaining three showed weak or variable activity. In contrast, neither of the TRL sites nor any of the five Su(Hw) sites we tested blocked enhancer-promoter interactions in this assay (Supplementary Fig. 22). These results support a role for CTCF in insulator activity *in vivo*, but indicate that other proteins that have classically been associated with insulator activity are not strictly linked to this function.

To further annotate predicted enhancers and to determine whether dynamics of chromatin and gene expression from whole animals can be associated with specific factors, we analysed the patterns of 38 diverse transcription factors we mapped at various developmental stages. We compared our data with the CAD database (Fig. 1e) and observed that many factors are specifically enriched in particular enhancer classes. For example, Engrailed (EN) binding sites are enriched within mesothoracic disc enhancers, whereas Knirps (KNI), Tailless and Schnurri (SHN) are enriched within blastoderm enhancers. Indeed, enhancers are usually characterized by multiple transcription factors binding in concert to target genomic DNA. We used a Gaussian kernel density estimation across the binding profiles of 38 transcription factors mapped in early embryos in this and two previously published studies<sup>7,9</sup> to define a 'transcription factor complexity' score based on the number and proximity of contributing transcription factors (see Supplementary Methods). Of 38,562 unique binding sites mapped by the 38 transcription factors, 38.3% are bound by more than two factors. Of the unique binding sites, 5.2% (1,962) are bound by more than eight factors (Supplementary Table 5 and Supplementary Figs 23, 24) and are considered high occupancy target (HOT) regions. Although HOT regions have been observed in *Caenorhabditis elegans*<sup>18</sup> and human (ENCODE project, unpublished results), their function in gene regulation is unknown<sup>7,19</sup>.

Regions of higher complexity are weakly associated with more highly expressed genes ( $r^2 = 0.19$ ), indicating that low-complexity binding sites are associated with more restricted expression patterns. Interestingly, annotated enhancers, CBP, activating histone marks including H3K4me1, and HDAC1, 4, 6 and 11 are most significantly enriched within low-to moderate- complexity category (CC) regions (CC2–CC8) (Fig. 2b). These enrichments consistently decrease at regions of high complexity (CC8–16). In contrast, we found that coding exons and HDAC3, which marks actively transcribed exons<sup>16</sup> (Fig. 2b and

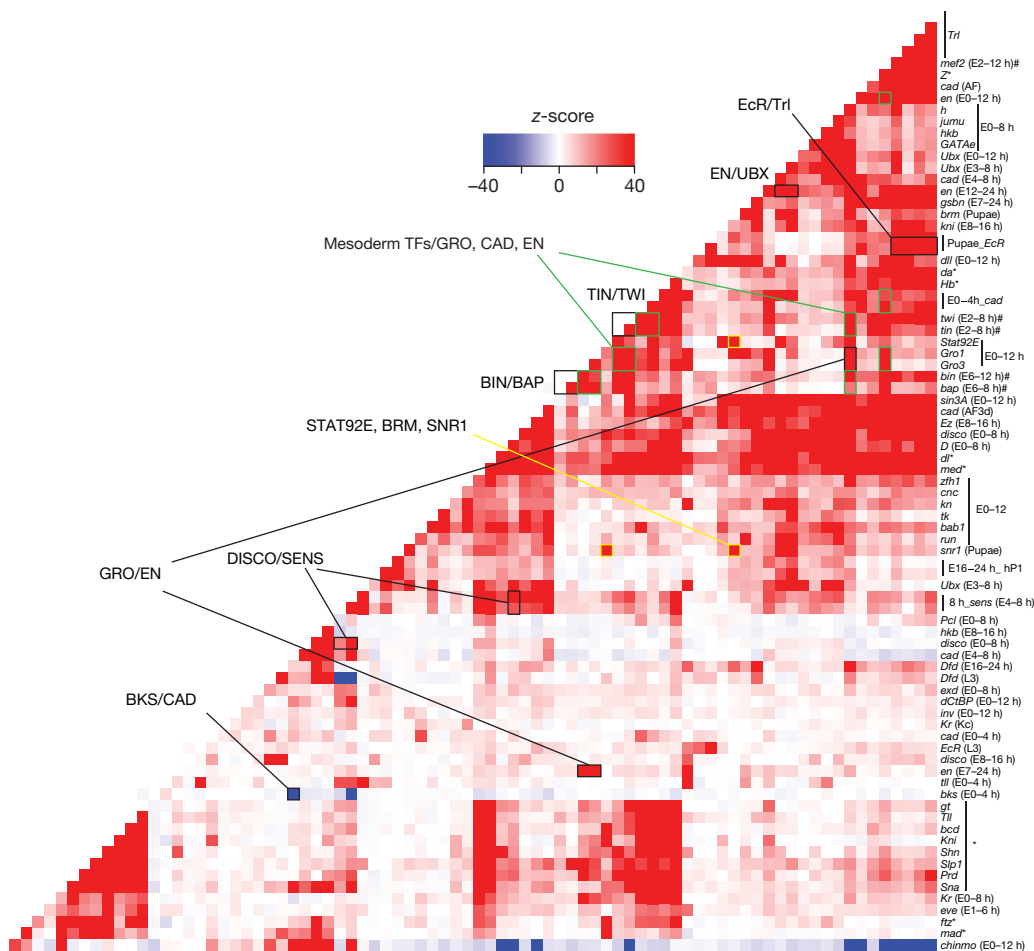
Supplementary Fig. 24), are depleted from moderate-to-high-complexity regions (>CC4). As expected, transcription factor complexity is inversely correlated with nucleosome enrichment<sup>20</sup> (Fig. 2b). Interestingly, when compared to our enhancer validations and negative controls that were selected independent of HOT region determination, there seems to be no obvious relationship between enhancer activity and HOT regions; 13 validated enhancers overlap with HOT regions but so did several sequences that give no enhancer activity (Fig. 1d, Supplementary Table 4 and data not shown). Taken together, these results indicate that HOT regions are primarily associated with open chromatin but that they do not always demarcate *cis*-regulatory elements.

The existence of HOT regions complicates the interpretation of transcription factor co-occurrence. For example, pair-wise clustering of TFBSs resulted in very large groups of co-occurring transcription factors, revealing few specific relationships (Supplementary Fig. 23). However, TFBS clustering performed on HOT-subtracted TFBSs reveals structure that is otherwise obscured when HOT regions are included (Fig. 3). For example, binding sites from different stages assayed for the same transcription factor (for example, TRL, Ultrabithorax (UBX), Ecdysone receptor (ECR)), known interactors (for example, Tinman (TIN) with Twist (TWI) and Biniou (BIN) with Bagpipe (BAP)), and technical replicates (for example, GRO) are more tightly clustered in the HOT-subtracted data. Transcription factors known to physically interact with one another at specific enhancers were significantly associated throughout the genome. For example, the co-repressor complex of GRO and EN and the *Drosophila* SWI/SNF chromatin remodelling complex components Brahma (BRM) and Snf5-related 1 (SNR1) show significant co-binding ( $z > 20$ ). Co-binding enrichment genome-wide was also observed for transcription factors that are known to bind independently to particular enhancers, such as UBX and EN that each bind to the DMX enhancer of the *distalless* (*dll*) gene, and each independently contribute to *dll* repression in different embryonic segments<sup>21</sup>. DLL was itself enriched for co-binding with EN, GRO and UBX, indicating common regulation of target genes. Interestingly, such previously undescribed interactions were seen at significance levels equal to or greater than those of known interactions. For example, while the previously reported mesodermal transcription factor data set<sup>9</sup> (TIN, TWI, BIN, BAM, MEF2) all have high overlap with one another as expected, these factors also all show highly significant overlap with GRO, CAD and EN. Many other notable overlap pairs were identified, including the Ecdysone receptor with TRL,



**Figure 2 | Transcription factor binding site complexity.** **a**, Number of TFBSs (left y-axis, black circles) and distribution of genomic annotation classes (right y-axis, colours) as a function of TFBS complexity (x-axis). **b**, Enrichment (colour scale: depleted in blue, enriched in red) of TFBSs sorted by complexity (x-axis) within annotated enhancers (CM, cardiac mesoderm; Ht, heart muscle; SM, somatic muscle; Meso., mesoderm; VM, visceral muscle), HDAC binding sites, early embryo chromatin marks. At the top is a heatmap depicting nucleosome density (Nuc. dens.) as a function of TFBS complexity.





**Figure 3 | Transcription factor binding site overlap.** Pairwise TFBS enrichment/depletion (colour-coded by z-score). TFBS data sets labelled at right. Asterisks indicate data from the BDTNP consortium<sup>6,7</sup>; hashes denote data from the Furlong laboratory<sup>9</sup>. Selected interactions described in the text are highlighted.

the peripheral nervous system master regulator Senseless (SENS)<sup>22</sup> with the axon guidance transcription factor Disconnected, and the Jak/Stat signalling pathway transcription factor STAT92E with the chromatin remodelling complex factors BRM and SNR1—all potential new connections between well-studied regulatory pathways or mechanisms. In total there are 831 highly significant positive pair-wise co-binding interactions in Supplementary Fig. 25 (z-score >20; bright red in Supplementary Fig. 25c), most of which are previously undescribed.

Although most significantly associated transcription factor pairs did show positive overlaps, we observed a few instances of highly significant negative associations (shown in blue, Fig. 3). One of the most anti-correlated pairs of transcription factors is Brakeless (BKS) and CAD. BKS is a co-repressor that has been implicated in gap gene regulation, for example acting to restrict the expression of *knirps* (*kni*) and *giant* (*gt*) in the posterior blastoderm<sup>23</sup>. In contrast, CAD activates *kni* and *gt* in the same embryonic domain<sup>24</sup>. Even when BKS and CAD have multiple binding sites nearby one another, they appear to be nonoverlapping and in different putative *cis*-regulatory elements (Supplementary Fig. 26). The biologically opposing roles of these two transcription factors seem to have led to the evolution of a very strong repulsion for occupying the same regulatory elements. To our knowledge, this genome-wide aversion in terms of transcription factor co-occupancy has not previously been observed in a metazoan genome.

To visualize the regulatory interactions among transcription factors, we built an intuitive hierarchy representing transcription factor regulatory associations (Supplementary Figs 25, 27; see Supplementary Methods). This network was constructed using 61 transcription factor data sets generated by the modENCODE project (pink nodes) and 20 transcription factors from recently published work<sup>6,7,9</sup> (green and yellow nodes). Specifically, we built a core hierarchy using a breadth-first search algorithm in a bottom-up fashion. Transcription factors

that regulate fewer than five transcription factors formed the bottom layer whereas transcription factors that directly regulated the bottom layer factors form the second layer. In total, the network model characterized 835 interactions; 686 were established by transcription factors mapped in this study (blue edges), 125 were derived from previously published data (grey edges), and 24 were auto-regulatory<sup>6,7,9</sup> (Supplementary Fig. 25). Components of the network derived from modENCODE-mapped transcription factors capture many known regulatory interactions; for example, EVE regulates *ftz* and *prd*. However, the vast majority of the 686 transcription factor interactions represent new putative regulatory relationships.

Transcription factors involved in widespread target co-binding and feed-forward loops are also likely to be involved in regulating common patterns of expression. To understand better how combinatorial transcription factor binding regulates developmentally dynamic gene expression, we analysed gene expression data from our RNA-seq time course and an independently performed 64-stage-developmental microarray expression time course. We partitioned the expression data sets into 18 and 64 *k*-means clusters, respectively, which resulted in gene sets with widely varying temporal specificity (Supplementary Fig. 25b). For each cluster of genes, we then quantified the enrichment of promoter-proximal binding sites for 90 modENCODE and previously published transcription factor data sets. From the microarray time-course clustering, five metaclusters were identified. Genes within these metaclusters are most highly expressed at third instar to adulthood (I), first instar to pupal-adult ecdysone pulse (II), early embryos (III), embryogenesis and larval life (IV) and late embryos (V). In both the microarray and RNA-seq time courses, most clusters are significantly associated with a core set of transcription factors including SIN3A, UBX, CAD, SENS and TRL. Interestingly, all metaclusters are enriched for TRL binding sites except V, which is enriched for SNR1, another Trithorax group gene; consistent

with reports that SNR1 has specialized functions<sup>25</sup>. Metacluster II is most highly expressed during adult central nervous system development<sup>26</sup> and enriched for several neuronal differentiation factors (Krüppel, KNI and Jumeau)<sup>27</sup>. Metacluster III uniquely is associated with embryonic patterning and organogenesis transcription factors (for example, Runt, Hunchback, TWI). Notably, many of the transcription factor co-enrichments within gene expression clusters correspond to binding site and regulatory co-enrichments (Fig. 3 and Supplementary Fig. 25), indicating that many of the co-associations of transcription factors with developmental expression patterns reflect co-binding and coordinate regulation at target sites in the genome.

In summary, we generated a draft regulatory annotation map of the *Drosophila* genome from 313 genome-wide data sets that identify or predict thousands of regulatory elements, including 537 silencers, 2,307 newly annotated promoters, 14,450 candidate CBP-bound *cis*-regulatory elements, 7,685 putative insulators<sup>10</sup> and 35,000 unique TBFSs that were bound by one or more transcription factor (Supplementary Tables 6–16 and Supplementary Fig. 28). The transcription factor binding results defined HOT regions of increased transcription factor complexity and their association with HDACs and open chromatin. Subsequent analysis of significantly co-bound transcription factors and transcription factor networks with HOT-subtracted data greatly expands the existing view of regulatory interactions among transcription factors and associates specific sets of transcription factors with specific developmental gene expression patterns. Several unexpected results arose from this initial phase of the modENCODE Project. For example, we revealed a specific class of unmarked promoters, identified a surprising association of HDAC4 and HDAC1/RPD3 to PREs, and discovered pairs of transcription factors that systematically avoid binding near each other throughout the genome. We expect that the results from modENCODE will serve as launch points for many new investigations, and that additional novel insights about the functional consequences of the patterns we describe here will emerge as others in the community engage with these data.

## METHODS SUMMARY

ChIP experiments were performed on whole *Drosophila melanogaster* animals from the following developmental stages: embryonic stages 0–4 h, 4–8 h, 12–16 h, 16–20 h, 20–24 h, larval stages L1, L2 and L3, pupal stage and adult male. The biological material was homogenized in 1.8% of formaldehyde. The cross-linked chromatin was sonicated to an average size of 500 bp. Pre-cleared chromatin extract was incubated overnight at 4 °C with the specific antibody and immunoprecipitated. ChIP material was hybridized either on custom Agilent tiling microarrays or on Affymetrix Tiling arrays. For ChIP-seq, standard protocols for Illumina Genome Analysers were used. The software packages used for peak detection were MACS, Peakseq, HGGSEG, CisGenome, MAT and HMMseq where appropriate. For RNA-seq experiments, total RNA was extracted from the same material used for ChIP and processed according to Illumina standard protocols. All methods and scripts used for the analysis of the data are described in Supplementary Information and are available on request. Transgenic assays for promoters, insulators and enhancers are described in Supplementary Information.

Received 23 September 2010; accepted 2 February 2011.

- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Johnson, D. S., Mortazavi, A., Myers, R. M. & Wold, B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* **316**, 1497–1502 (2007).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
- Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**, 212–224 (2010).
- Li, X. Y. *et al.* Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27 (2008).
- MacArthur, S. *et al.* Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* **10**, R80 (2009).
- Zeitlinger, J. *et al.* RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nature Genet.* **39**, 1512–1516 (2007).

- Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E. E. Combinatorial binding predicts spatio-temporal *cis*-regulatory activity. *Nature* **462**, 65–70 (2009).
- Nègre, N. *et al.* A comprehensive map of insulator elements for the *Drosophila* genome. *PLoS Genet.* **6**, e1000814 (2010).
- Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
- Schotta, G. *et al.* Central role of *Drosophila* SU(VAR)3–9 in histone H3-K9 methylation and heterochromatic gene silencing. *EMBO J.* **21**, 1121–1131 (2002).
- Agger, K., Christensen, J., Cloos, P. A. & Helin, K. The emerging functions of histone demethylases. *Curr. Opin. Genet. Dev.* **18**, 159–168 (2008).
- Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).
- Kwong, C. *et al.* Stability and dynamics of polycomb target sites in *Drosophila* development. *PLoS Genet.* **4**, e1000178 (2008).
- Kolasinska-Zwier, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature Genet.* **41**, 376–381 (2009).
- Hoskins, R. A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192 (2011).
- Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
- Moorman, C. *et al.* Hotspots of transcription factor colocalization in the genome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **103**, 12027–12032 (2006).
- Henikoff, S., Henikoff, J. G., Sakai, A., Loeb, G. B. & Ahmad, K. Genome-wide profiling of salt fractions maps physical properties of chromatin. *Genome Res.* **19**, 460–469 (2009).
- Gebelein, B., McKay, D. J. & Mann, R. S. Direct integration of *Hox* and segmentation gene inputs during *Drosophila* development. *Nature* **431**, 653–659 (2004).
- Nolo, R., Abbott, L. A. & Bellen, H. J. Senseless, a Zn finger transcription factor, is necessary and sufficient for sensory organ development in *Drosophila*. *Cell* **102**, 349–362 (2000).
- Haeccker, A. *et al.* *Drosophila* brakeless interacts with atrophin and is required for tailless-mediated transcriptional repression in early embryos. *PLoS Biol.* **5**, e145 (2007).
- Rivera-Pomar, R., Lu, X., Perrimon, N., Taubert, H. & Jackle, H. Activation of posterior gap gene expression in the *Drosophila* blastoderm. *Nature* **376**, 253–256 (1995).
- Zrally, C. B. *et al.* SNR1 is an essential subunit in a subset of *Drosophila* brm complexes, targeting specific functions during development. *Dev. Biol.* **253**, 291–308 (2003).
- Truman, J. W. Metamorphosis of the central nervous system of *Drosophila*. *J. Neurobiol.* **21**, 1072–1084 (1990).
- Parrish, J. Z., Kim, M. D., Jan, L. Y. & Jan, Y. N. Genome-wide analyses identify transcription factors required for proper morphogenesis of *Drosophila* sensory neuron dendrites. *Genes Dev.* **20**, 820–835 (2006).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by U01HG004264 from the National Human Genome Research Institute to K.P.W. and also funded by the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Human Genome Research Institute (NHGRI) or the National Institutes of Health (NIH). C.D.B. is supported by a Lilly-Life Sciences Research Foundation fellowship. C.A.B. is supported by a NIH NRSA postdoctoral fellowship. R.P.A. is in part supported by an Isaac Newton Trust award to R.W. P.L. was supported by a grant from the Department of Energy Computational Sciences Graduate Fellowship (DOE CSGF). M.E.L. and D.M.M. work was supported by NHGRI grant U01 HG004279. We thank the Functional Genomics Facility at the University of Chicago and the High-Throughput Genome Analysis Core at Argonne National Laboratory for processing of microarrays and of Illumina sequence. We thank T.-R. Li, J. D. Lambert, S. Rifkin, T. Herreman, C. Mason, L. Sun and Z. Gauhar for producing the developmental expression microarray data. We also thank the many members of the *Drosophila* community who contributed to this work by providing reagents. A complete list of community participants is included in the Supplementary Methods.

**Author Contributions** N.N., L.S. and K.P.W. designed and produced modENCODE antibodies; N.N., Z.L., H.I., R.F.S., M.B.D., C.A.M., J.Z., S.S. and M.D. performed the ChIP-chip and ChIP-seq experiments; R.F.S., K.V., H.B. and A.V. produced the GFP-tagged transcription factor *Drosophila* lines; S.W.M., H.I., L.H. and R.P.A. performed the validation experiments of promoters, enhancers and insulators; N.N., P.K.S., N.A.B., A.J.G., D.H. and R.L.G. performed the primary analysis and organized the data of ChIP-chip and ChIP-seq experiments; N.N., C.D.B., L.M., C.A.B., U.W., P.K., M.E.L., P.L., R.S., J.C., C.C., P.K.S., D.M.M. and M.G. analysed the data; M.M. contributed to reagents; N.N., C.D.B., L.M., C.A.B., S.W.M., R.P.A., R.W., S.R., B.R., M.G., J.W.P., M.K. and K.P.W. wrote the paper; H.B., R.W., S.R. (silencer/insulator analysis), R.L.G. (informatics), B.R. (chromatin data and promoter validation), J.W.P. (enhancer/promoter validation), M.K. (data analysis) and K.P.W. (project director) supervised the work.

**Author Information** Data were deposited at GEO under accession numbers GSE23537, GSE115292, GSE20000, GSE16245, GSE25955, GSE25964, GSE25956, GSE25957, GSE25958, GSE25959, GSE25960, GSE25961, GSE25962 and GSE25963. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to K.P.W. ([kpwhite@uchicago.edu](mailto:kpwhite@uchicago.edu)).

# TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA

Rachel Litman Flynn<sup>1</sup>, Richard C. Centore<sup>1\*</sup>, Roderick J. O'Sullivan<sup>3\*</sup>, Rekha Rai<sup>4\*</sup>, Alice Tse<sup>1</sup>, Zhou Songyang<sup>5</sup>, Sandy Chang<sup>4</sup>, Jan Karlseder<sup>3</sup> & Lee Zou<sup>1,2</sup>

**Maintenance of telomeres requires both DNA replication and telomere 'capping' by shelterin. These two processes use two single-stranded DNA (ssDNA)-binding proteins, replication protein A (RPA) and protection of telomeres 1 (POT1). Although RPA and POT1 each have a critical role at telomeres, how they function in concert is not clear. POT1 ablation leads to activation of the ataxia telangiectasia and Rad3-related (ATR) checkpoint kinase at telomeres<sup>1,2</sup>, suggesting that POT1 antagonizes RPA binding to telomeric ssDNA. Unexpectedly, we found that purified POT1 and its functional partner TPP1 are unable to prevent RPA binding to telomeric ssDNA efficiently. In cell extracts, we identified a novel activity that specifically displaces RPA, but not POT1, from telomeric ssDNA. Using purified protein, here we show that the heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1) recapitulates the RPA displacing activity. The RPA displacing activity is inhibited by the telomeric repeat-containing RNA (TERRA) in early S phase, but is then unleashed in late S phase when TERRA levels decline at telomeres<sup>3</sup>. Interestingly, TERRA also promotes POT1 binding to telomeric ssDNA by removing hnRNPA1, suggesting that the reaccumulation of TERRA after S phase helps to complete the RPA-to-POT1 switch on telomeric ssDNA. Together, our data suggest that hnRNPA1, TERRA and POT1 act in concert to displace RPA from telomeric ssDNA after DNA replication, and promote telomere capping to preserve genomic integrity.**

RPA binds ssDNA in a non-sequence specific manner<sup>4</sup>, whereas POT1 specifically recognizes ssDNA consisting of the telomeric repeats<sup>5</sup>. RPA plays a key role in DNA replication and activation of the ATR checkpoint<sup>6</sup>, and POT1 suppresses ATR activation at telomeres<sup>1,2</sup> (Supplementary Fig. 1). In both yeast and humans, RPA associates with telomeres during S phase of the cell cycle<sup>7–9</sup>, and is implicated in telomere maintenance<sup>10–12</sup>. Furthermore, ATR transiently associates with telomeres and suppresses telomere instability<sup>7,10,13</sup>. These findings raise the question of how the bindings of POT1 and RPA to telomeric ssDNA are orchestrated and, furthermore, how the interplay between POT1 and RPA affects DNA replication and ATR activation at telomeres.

Double-stranded DNA (dsDNA) with ssDNA overhangs activates ATR in human cell extracts<sup>14</sup>. To investigate how ATR activation is suppressed at telomeres, we tested whether telomeric ssDNA overhangs affect ATR activation in this assay. Resected dsDNA of random sequences, but not resected telomeric dsDNA, efficiently induced the phosphorylation of RPA2 by ATR (Supplementary Fig. 2)<sup>14</sup>, suggesting that telomeric ssDNA overhangs do not support efficient ATR activation in cell extracts.

The absence of ATR activation by telomeric ssDNA suggests that POT1 may prevent RPA binding to telomeric ssDNA<sup>2</sup>. POT1 and TPP1 function as heterodimers in cells, and the complex binds to telomeric ssDNA more efficiently than POT1 alone<sup>15,16</sup>. In gel-shift assays, the POT1–TPP1 complexes purified from insect or human cells

and the RPA purified from *Escherichia coli* efficiently bound to a telomeric ssDNA probe (Fig. 1a and Supplementary Fig. 3a, b). POT1–TPP1 exhibited lower affinity for telomeric ssDNA than RPA (Supplementary Fig. 3a). When POT1–TPP1 and RPA were co-incubated with the probe, the RPA–ssDNA complex was readily detected, whereas no POT1-containing complexes were observed (Fig. 1a and Supplementary Fig. 3b). In pull-down assays using biotinylated telomeric ssDNA (ssTEL), RPA also outcompeted POT1–TPP1 for binding to ssTEL (Fig. 1b and Supplementary Fig. 3c). Thus, RPA, which is more abundant than POT1–TPP1 in cells<sup>4,17</sup>, outcompetes POT1–TPP1 for binding to telomeric ssDNA when present at similar concentrations as POT1–TPP1. The *E. coli* ssDNA-binding protein only modestly reduced POT1 binding to ssTEL (Supplementary Fig. 3c), suggesting that the ability to outcompete POT1–TPP1 is unique to RPA.

The ability of RPA to outcompete POT1–TPP1 raises the question of how ATR activation is suppressed in cell extracts. Purified RPA bound to ssTEL and mutated telomeric repeats (ssMUT) efficiently (Fig. 1c). In stark contrast to purified RPA, the endogenous RPA in HeLa whole-cell extracts (WCEs) was largely excluded from ssTEL; however, it still associated with ssMUT (Fig. 1c). The sequence-specific exclusion of RPA from ssTEL in WCEs suggests that RPA may be outcompeted by other proteins or actively displaced from telomeric ssDNA.

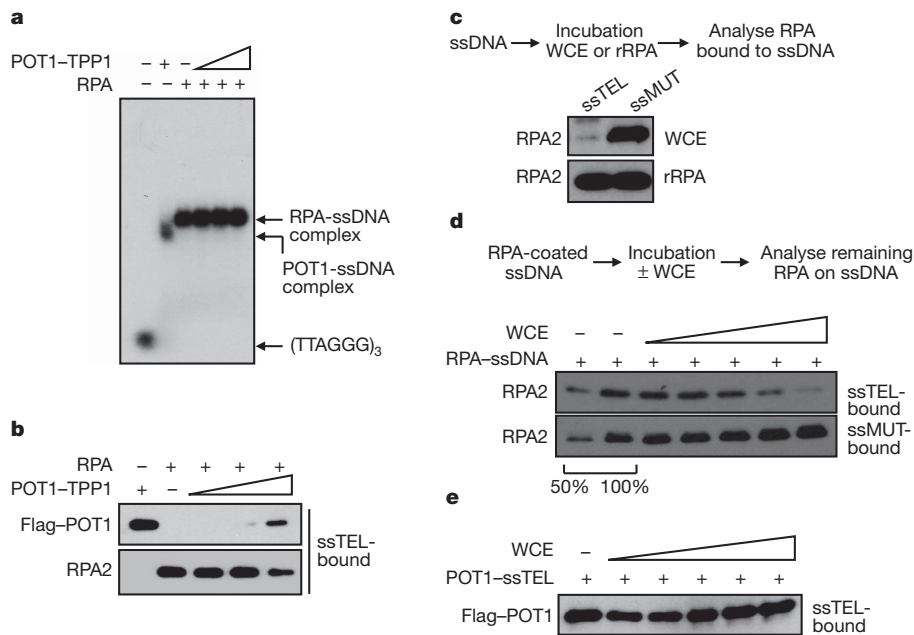
To assess if RPA is actively displaced from ssTEL, we pre-coated ssTEL and ssMUT with RPA then incubated them in extracts. The levels of RPA on ssTEL gradually declined with increasing concentrations of WCEs from HeLa, HEK293E, U2OS and MEF cells (Fig. 1d and Supplementary Fig. 4a). In addition, HeLa nuclear extracts, but not the cytoplasmic extracts, efficiently displaced RPA from ssTEL (Supplementary Fig. 4b). In marked contrast to the RPA on ssTEL, the RPA bound to ssMUT remained constant regardless of WCE concentrations (Fig. 1d). When POT1-coated ssTEL was incubated in extracts, POT1 remained stably bound to ssTEL even in high concentrations of WCEs (Fig. 1e). Furthermore, RPA was rapidly displaced from ssTEL within 5 min, whereas no POT1 was displaced after 60 min (Supplementary Fig. 4c). Thus, the activity that displaces RPA from telomeric ssDNA is sequence-specific, protein-specific and localized within the nucleus.

The specific displacement of RPA, but not POT1, from telomeric ssDNA prompted us to test if POT1 is the RPA displacing factor. When incubated with RPA–ssTEL, POT1–TPP1 did not significantly reduce the levels of ssTEL-bound RPA (Supplementary Fig. 5a). To identify the RPA displacing factors, we sought to capture the RPA displacing activity from extracts using RPA–ssTEL as bait. The proteins captured and eluted from RPA–ssTEL, but not RPA–ssMUT, recapitulated the RPA displacing activity (Fig. 2a). Mass spectrometry analysis of the proteins specifically captured by RPA–ssTEL identified hnRNPA1

<sup>1</sup>Massachusetts General Hospital Cancer Center, Harvard Medical School, Charlestown, Massachusetts 02129, USA. <sup>2</sup>Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Molecular and Cellular Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA. <sup>4</sup>Department of Laboratory Medicine, Yale University School of Medicine, New Haven, Connecticut 06510, USA. <sup>5</sup>Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas 77030, USA.

\*These authors contribute equally to this work.





**Figure 1 | A novel telomere-specific RPA displacing activity in human cell extracts.** **a**, POT1–TPP1 (60 nM; purified from insect cells), RPA (60 nM) and mixtures of POT1–TPP1 and RPA (60, 120, 180 nM of POT1–TPP1 mixed with 60 nM of RPA) were incubated with 20 nM of the ssDNA probe and analysed by gel-shift. **b**, POT1–TPP1 (2.4 nM), RPA (2.4 nM) and mixtures of POT1–TPP1 and RPA (2.4, 4.8, 7.2 nM of POT1–TPP1 mixed with 2.4 nM of RPA) were incubated with 0.8 nM of biotinylated ssTEL ((TTAGGG)<sub>8</sub>). The

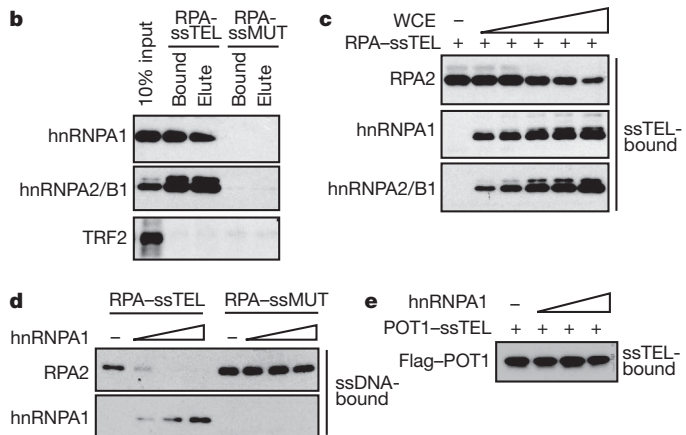
and hnRNPA2/B1 (Supplementary Fig. 5b, c), both of which are known to bind telomeric ssDNA<sup>18,19</sup>. The presence of hnRNPA1 and A2/B1 in the eluted fraction with RPA displacing activity was confirmed by western blot (Fig. 2b). Moreover, hnRNPA1 and A2/B1 gradually bound to ssTEL as RPA was displaced in WCEs (Fig. 2c). These results suggest that hnRNPA1 and A2/B1 may play a role in RPA displacement.

hnRNPA1 has been implicated in telomere maintenance<sup>20,21</sup>. Extracts from hnRNPA1 knockdown cells exhibited reduced activity

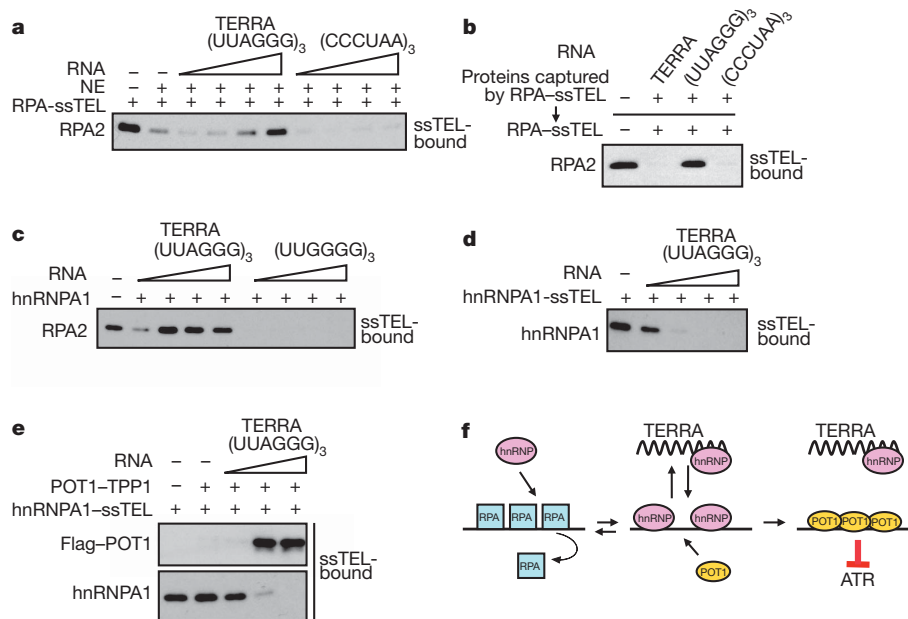
proteins bound to ssTEL were retrieved by streptavidin beads and analysed by western blot. **c**, Biotinylated ssTEL or ssMUT ((TTTGGCG)<sub>8</sub>) was incubated with WCEs or recombinant RPA (rRPA). **d**, ssTEL or ssMUT pre-coated with RPA was incubated with increasing concentrations of HeLa WCEs (0.08, 0.19, 0.36, 0.8, 1.3 µg µl<sup>-1</sup>). The RPA2 remaining on ssTEL was analysed as in **b**. **e**, ssTEL pre-coated with POT1 was incubated with increasing concentrations of HeLa WCEs (0.07, 0.18, 0.33, 0.66, 1.3 µg µl<sup>-1</sup>).

in RPA displacement (Supplementary Fig. 6a). Purified hnRNPA1 efficiently displaces RPA from ssTEL, but not ssMUT (Fig. 2d). Furthermore, hnRNPA1 did not displace POT1 from ssTEL (Fig. 2e). hnRNPA1 only displaces RPA from ssTEL containing four or more telomeric repeats (Supplementary Fig. 6b), indicating that a DNA length-dependent binding mode of hnRNPA1 may be needed to displace RPA<sup>22</sup>. Given that hnRNPA1 and A2/B1 are highly homologous in the RRM domains that bind telomeric ssDNA, both of these hnRNPs may contribute to RPA displacement.

hnRNPA1 not only binds telomeric ssDNA but also TERRA<sup>23–26</sup>. To test if TERRA affects the ability of hnRNPA1 to displace RPA from ssTEL, we added increasing concentrations of TERRA or control RNA to nuclear extracts, then incubated the extracts with RPA–ssTEL. RPA displacement was virtually abolished by TERRA, but not control RNA (Fig. 3a). The RPA displacing activity captured by RPA–ssTEL was also specifically inhibited by TERRA (Fig. 3b). Furthermore, the ability of purified hnRNPA1 to bind ssTEL and to displace RPA from ssTEL was specifically inhibited by TERRA (Fig. 3c and Supplementary Fig. 6c). Thus, TERRA is a potent inhibitor of the RPA displacing activity of hnRNPA1.



**Figure 2 | RPA displacement by hnRNPA1.** **a**, ssTEL and ssMUT pre-coated with RPA were incubated with nuclear extracts (NE). After the incubation, the proteins bound to DNA were retrieved, eluted and applied to RPA-coated ssTEL or ssMUT (see Supplementary Methods). After the second incubation, the remaining RPA2 on DNA was analysed by western blot. **b**, Proteins captured by RPA–ssTEL or RPA–ssMUT and eluted by salt were analysed by western blot using antibodies to hnRNPA1, hnRNPA2/B1 and TRF2. **c**, RPA-coated ssTEL (0.8 nM) was incubated with increasing concentrations of WCES (0.06, 0.24, 0.96  $\mu\text{g } \mu\text{l}^{-1}$ ). The hnRNPA1 and hnRNPA2/B1 bound to DNA and the remaining RPA2 on DNA were analysed by western blot. **d**, RPA-coated ssTEL or ssMUT (0.8 nM) was incubated with increasing concentrations of purified hnRNPA1 (2.4, 4.8, 7.2 nM). The remaining RPA2 on DNA was analysed as in **a**. **e**, POT1-coated ssTEL (0.8 nM) was incubated with increasing concentrations of purified hnRNPA1 (2.4, 4.8, 7.2 nM).



**Figure 3 | Regulation of RPA displacement by TERRA.** **a**, Nuclear extracts (34 ng  $\mu\text{l}^{-1}$ ) were treated with increasing concentrations (2, 4, 10, 20 nM) of TERRA (UUAGGG)<sub>3</sub>, control RNA (CCCUGA)<sub>3</sub> or mock treated. The treated nuclear extracts were then incubated with RPA-coated ssTEL (2 nM), and the remaining RPA2 on ssTEL was analysed after the incubation. **b**, The RPA displacing factors were captured with RPA-ssTEL as in Fig. 2a. The elution was incubated with TERRA or control RNA, then applied to RPA-ssTEL. **c**, Purified hnRNPA1 (4.8 nM) was incubated with increasing concentrations of

If hnRNPA1 displaces RPA from telomeric ssDNA, how can POT1 bind to telomeric ssDNA? Given that hnRNPA1 has affinity for both telomeric ssDNA and TERRA, the presence of TERRA at telomeres may promote the dissociation of hnRNPA1 from telomeric ssDNA. Indeed, when hnRNPA1-coated ssTEL was incubated with TERRA, hnRNPA1 was stripped from ssTEL (Fig. 3d), showing that hnRNPA1 binds telomeric ssDNA dynamically. Furthermore, when hnRNPA1-ssTEL was incubated with TERRA and POT1-TPP1, POT1 efficiently bound to ssTEL as hnRNPA1 was removed by TERRA (Fig. 3e).

The *in vitro* results above suggest that the initial displacement of RPA from telomeric ssDNA may be performed by hnRNPs when TERRA levels are low at telomeres (Fig. 3f). However, if TERRA levels rise at telomeres, hnRNPA1 may shuttle between telomeric ssDNA and TERRA dynamically. In this situation, both RPA and POT1 may have the chance to bind telomeric ssDNA. Because hnRNPA1 only displaces RPA, but not POT1, this dynamic process will eventually promote POT1 occupancy at telomeric ssDNA.

This model raises the possibility that the RPA displacing activity may be regulated by TERRA during the cell cycle. To test this, we generated WCEs from cells in G1, early S, late S and M phases of the cell cycle. RPA was more efficiently displaced in the late S- and M-phase extracts than in the G1- and early S-phase extracts (Fig. 4a and Supplementary Fig. 7a, b). Thus, the RPA displacing activity is low in G1 and early S phase, but upregulated in late S phase.

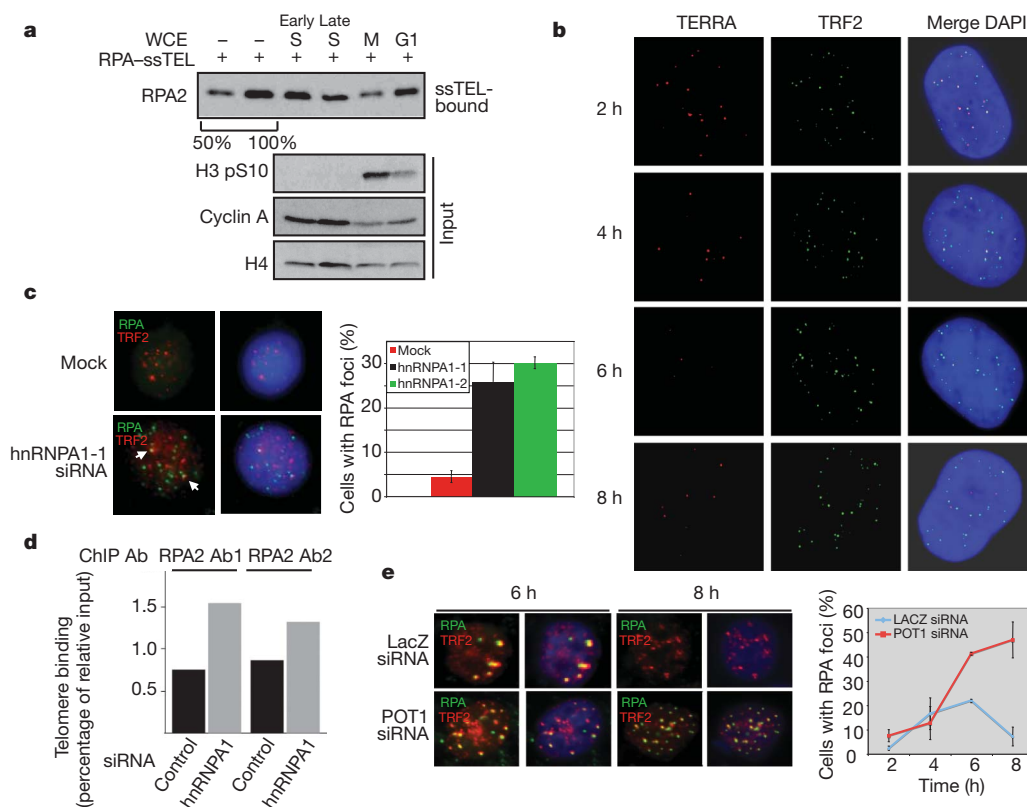
If TERRA inhibits the RPA displacing activity, its levels should inversely correlate with the activity. Furthermore, removal of TERRA in early S phase should alleviate the inhibition. Indeed, a recent study showed that TERRA levels significantly decrease in late S phase and increase again after S phase<sup>3</sup>. Consistently, telomeric TERRA foci declined as cells progressed from early to late S phase (Fig. 4b and Supplementary Fig. 7c, d). In addition, RNase A treatment of early S-phase extracts significantly enhanced the RPA displacing activity (Supplementary Fig. 7e). Together, these results suggest that TERRA inhibits RPA displacement in early S phase, and its decline in late S phase may provide a window for RPA displacement.

TERRA or control RNA (2, 4, 10, 20 nM), then incubated with RPA-ssTEL (0.8 nM). **d**, hnRNPA1-coated ssTEL (0.8 nM) was incubated with increasing concentrations of TERRA (2, 20, 200, 2,000 nM). The remaining hnRNPA1 on ssTEL was analysed by western blot. **e**, hnRNPA1-coated ssTEL (0.8 nM) was incubated with increasing concentrations of TERRA (2, 20, 200 nM) in the presence of POT1-TPP1 (2.4 nM). The hnRNPA1 and POT1 on ssTEL were analysed by western blot. **f**, A model for RPA displacement.

The model above also predicts that hnRNPs are necessary for RPA displacement from telomeres. Depletion of hnRNPA1 using two independent short interfering RNA (siRNAs) significantly increased the fraction of cells displaying RPA foci (Fig. 4c and Supplementary Fig. 8a–d). Notably, a fraction of the RPA foci in hnRNPA1 knockdown cells closely associated with TRF2 foci. Furthermore, increased RPA binding at telomeres was detected in hnRNPA1 knockdown cells by chromatin immunoprecipitation (Fig. 4d). In synchronized hnRNPA1 knockdown cells, RPA binding to telomeres was enhanced in early S phase (Supplementary Fig. 9a, b), indicating that even during this period some hnRNPA1 remains free from TERRA and limits RPA binding to telomeres<sup>9</sup>. In late S/G2, RPA still declined at telomeres in hnRNPA1 knockdown cells, possibly owing to the redundancy among hnRNPs.

If the displacement of RPA by hnRNPA1 is a prerequisite for POT1 binding, POT1 should be needed for RPA exclusion after late S phase. To assess this possibility, we treated cells with POT1 siRNA and synchronized the cells with thymidine as POT1 levels declined (Supplementary Fig. 10a, b). After POT1 knockdown cells and control cells were synchronously released, RPA foci appeared in both cell populations (Fig. 4e). As control cells entered G2, RPA foci rapidly declined. In contrast, the POT1 knockdown cells containing RPA foci that colocalized with TRF2 continued to increase. Concomitantly, modest Chk1 phosphorylation was detected in POT1 knockdown cells (Supplementary Fig. 10c). Thus, reduction of POT1 compromises the exclusion of RPA from telomeres after replication<sup>27</sup>.

During early to middle S phase, TERRA sequesters hnRNPs and allows RPA to bind telomeric ssDNA at replication forks or telomere ends (Supplementary Fig. 1). When TERRA levels decline in late S phase, hnRNPs are unleashed to displace RPA from telomeric ssDNA. The dynamic binding of hnRNPs to telomeric ssDNA is gradually antagonized by TERRA when TERRA reaccumulates at telomeres, providing a window for both RPA and POT1 to bind. Because only POT1, but not RPA, binds to telomeric ssDNA irreversibly in the presence of hnRNPs, this dynamic process favours the formation of



**Figure 4 | hnRNP A1 and POT1 suppress the accumulation of RPA at telomeres.** **a**, RPA-coated ssTEL was incubated with WCEs from cells in G1, early S, late S and M phases of the cell cycle (see Supplementary Methods). The remaining RPA2 on ssTEL was analysed after incubation. Cyclin A and phospho-histone H3 serve as cell-cycle markers, and histone H4 as a loading control. **b**, TERRA was analysed by RNA fluorescence *in situ* hybridization in HeLa cells after thymidine release. TRF2 serves as a marker of telomeres. **c**, HeLa cells were treated with hnRNP A1 siRNA or mock treated, then immunostained with antibodies to RPA2 and TRF2 (left panel). The cells with

RPA foci ( $>5$ ) were quantified (right panel). Mean  $\pm$  s.d.,  $n = 3$  for mock and sihnRNP A1-1,  $n = 2$  for sihnRNP A1-2. **d**, Chromatin immunoprecipitation (ChIP) of RPA was performed with two different RPA2 antibodies (Ab). The association of RPA with telomeres was analysed by dot blot using a telomere probe and quantified. **e**, HeLa cells transfected with POT1 or LacZ siRNA were released from a thymidine block. At the indicated times, the G2/M population was determined by FACS (Supplementary Fig. 10c). Cells were immunostained for RPA2 and TRF2 (left panel). The cells with RPA foci ( $>5$ ) were quantified (right panel). Mean  $\pm$  s.d.,  $n = 3$ .

POT1-coated telomeric ssDNA. Unlike RPA, POT1 kinks telomeric ssDNA and induces its self-recognition<sup>28,29</sup>. These unique properties of POT1 may confer resistance to hnRNP-mediated displacement. The cell-cycle-regulated RPA displacement may allow RPA to transiently associate with telomeric ssDNA during replication, and prevent persistent ATR activation at telomeres after S phase. Once coated by POT1, telomeric ssDNA may remain capped until the arrival of replication forks in the next S phase. Together, TERRA and hnRNPs orchestrate a cell-cycle-regulated RPA-to-POT1 switch on telomeric ssDNA, ensuring orderly telomere replication and capping.

## METHODS SUMMARY

To analyse the bindings of RPA and POT1-TPP1 to ssDNA, biotinylated ssDNA was attached to streptavidin-coated magnetic beads. Biotinylated ssDNA (1 pmol) was incubated with purified protein in 500  $\mu$ l of binding buffer (10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10  $\mu$ g ml<sup>-1</sup> BSA, 10% glycerol, 0.05% NP-40).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 8 August; accepted 22 December 2010.

Published online 13 March 2011.

- Guo, X. *et al.* Dysfunctional telomeres activate an ATM-ATR-dependent DNA damage response to suppress tumorigenesis. *EMBO J.* **26**, 4709–4719 (2007).
- Denchi, E. L. & de Lange, T. Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* **448**, 1068–1071 (2007).
- Porro, A., Feuerhahn, S., Reichenbach, P. & Lingner, J. Molecular dissection of TERRA biogenesis unveils the presence of distinct and multiple regulatory pathways. *Mol. Cell Biol.* **30**, 4808–4817 (2010).

- Wold, M. S. Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu. Rev. Biochem.* **66**, 61–92 (1997).
- Baumann, P. & Cech, T. R. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* **292**, 1171–1175 (2001).
- Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
- Verdun, R. E. & Karlseder, J. The DNA damage machinery and homologous recombination pathway act consecutively to protect human telomeres. *Cell* **127**, 709–720 (2006).
- Moser, B. A. *et al.* Differential arrival of leading and lagging strand DNA polymerases at fission yeast telomeres. *EMBO J.* **28**, 810–820 (2009).
- McGee, J. S. *et al.* Reduced Rif2 and lack of Mec1 target short telomeres for elongation rather than double-strand break repair. *Nature Struct. Mol. Biol.* **17**, 1438–1445 (2010).
- Verdun, R. E., Crabbe, L., Haggblom, C. & Karlseder, J. Functional human telomeres are recognized as DNA damage in G2 of the cell cycle. *Mol. Cell* **20**, 551–561 (2005).
- Kibe, T., Ono, Y., Sato, K. & Ueno, M. Fission yeast Taz1 and RPA are synergistically required to prevent rapid telomere loss. *Mol. Biol. Cell* **18**, 2378–2387 (2007).
- Schramke, V. *et al.* RPA regulates telomerase action by providing Est1p access to chromosome ends. *Nature Genet.* **36**, 46–54 (2004).
- McNees, C. J. *et al.* ATR suppresses telomere fragility and recombination but is dispensable for elongation of short telomeres by telomerase. *J. Cell Biol.* **188**, 639–652 (2010).
- Shiotani, B. & Zou, L. Single-stranded DNA orchestrates an ATM-to-ATR switch at DNA breaks. *Mol. Cell* **33**, 547–558 (2009).
- Wang, F. *et al.* The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* **445**, 506–510 (2007).
- Xin, H. *et al.* TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* **445**, 559–562 (2007).
- Takai, K. K., Hooper, S., Blackwood, S., Gandhi, R. & de Lange, T. *In vivo* stoichiometry of shelterin components. *J. Biol. Chem.* **285**, 1457–1467 (2010).
- Ishikawa, F., Matunis, M. J., Dreyfuss, G. & Cech, T. R. Nuclear proteins that bind the pre-mRNA 3' splice site sequence (UUG/G) and the human telomeric DNA sequence d(TTAGGG)<sub>n</sub>. *Mol. Cell Biol.* **13**, 4301–4310 (1993).



19. McKay, S. J. & Cooke, H. hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. *Nucleic Acids Res.* **20**, 6461–6464 (1992).
20. LaBranche, H. *et al.* Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nature Genet.* **19**, 199–202 (1998).
21. Zhang, Q. S., Manche, L., Xu, R. M. & Krainer, A. R. hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**, 1116–1128 (2006).
22. Ding, J. *et al.* Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115 (1999).
23. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol. Cell* **35**, 403–413 (2009).
24. de Silanes, I. L., d'Alcontres, M. S. & Blasco, M. A. TERRA transcripts are bound by a complex array of RNA-binding proteins. *Nature Commun.* **1**, 1–9 (2010).
25. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**, 798–801 (2007).
26. Redon, S., Reichenbach, P. & Lingner, J. The non-coding RNA TERRA is a natural ligand and direct inhibitor of human telomerase. *Nucleic Acids Res.* **38**, 5797–5806 (2010).
27. Gong, Y. & de Lange, T. A. Shld1-controlled POT1a provides support for repression of ATR signaling at telomeres through RPA exclusion. *Mol. Cell* **40**, 377–387 (2010).
28. Lei, M., Podell, E. R. & Cech, T. R. Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nature Struct. Mol. Biol.* **11**, 1223–1229 (2004).
29. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank T. de Lange, A. Krainer, B. Chabot and M. Wold for reagents, and members of the Zou laboratory for discussion. L.Z. is an Ellison New Scholar on Aging. R.L.F. is supported by National Institutes of Health (NIH) fellowship 5T32CA009216-28 and American Cancer Society fellowship 0902501. R.C.C. is supported by NIH fellowship F32-GM089150. R.J.O. is supported by the George E. Hewitt Foundation for Medical Research. This work is supported by a Welch Foundation grant (Q-1673) to Z.S., an ACS grant (RSG-08-297) to L.Z., and NIH grants CA133249 to Z.S., CA129037 to S.C., GM06525 and AG025837 to J.K. and GM076388 to L.Z.

**Author Contributions** R.L.F. and L.Z. conceived the project. R.L.F., R.C.C., R.J.O., R.R., A.T. and L.Z. performed the experiments. Z.S. contributed the POT1–TPP1 complex purified from insect cells. R.R. and S.C. performed the combined RNA-fluorescence *in situ* hybridization and immunostaining analysis. R.J.O. and J.K. performed the RPA chromatin immunoprecipitation and northern blot. R.L.F. and L.Z. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to L.Z. ([zou.lee@mgh.harvard.edu](mailto:zou.lee@mgh.harvard.edu)).

## METHODS

**ATR activation.** To generate the 800-base-pair (bp) telomeric dsDNA fragment, the pSTY11 plasmid (a gift from T. de Lange) was digested with EcoRI and the excised fragment was gel purified. The 800-bp random sequence dsDNA was generated by PCR and column purified. These dsDNA fragments were incubated with T7 exonuclease for 15 s at room temperature (approximately 23 °C) and flash frozen in an ethanol–dry-ice bath. T7 was inactivated by subsequent incubation at 70 °C for 20 min and DNA fragments were separated on 2% agarose gel to confirm equal resection. The resected DNA fragments were incubated with nuclear extract as previously described<sup>14</sup>. To monitor specifically the phosphorylation of RPA2 by ATR and eliminate the contributions of ataxia telangiectasia mutated (ATM) and DNA-PK to RPA2 phosphorylation, nuclear extracts were pre-treated with 20  $\mu$ M KU55933 and NU7026 inhibitors for 15 min at 4 °C. The extracts were mixed with the DNA fragments, incubated for 15 min at 37 °C, and RPA phosphorylation was analysed by western blot.

**Protein purification.** The POT1–TPP1 complex was either purified from baculovirus-infected Sf9 cells as previously described<sup>16</sup>, or purified from HEK293E cells as follows. The pCL–Flag–POT1 and pCL–Flag–TPP1 vectors<sup>16</sup> were individually transfected or co-transfected into HEK293E cells. The cells were collected after 72 h and lysed in the NETN buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris–HCl (pH 8.0), 0.5% NP-40 and protease inhibitors), sonicated and cleared by centrifugation (10,000g for 10 min). The cleared lysates were incubated with the M2 anti-Flag antibody-conjugated beads at 4 °C for 2 h and eluted with 200  $\mu$ g ml<sup>−1</sup> 3 $\times$  Flag peptide in buffer A (25 mM Tris–HCl (pH 8.0) 100 mM NaCl, 10% glycerol) for 1 h. Recombinant RPA complex was purified from *E. coli* as previously described<sup>30</sup>. hnRNPA1 pET9d plasmid (a gift from A. Krainer) was transformed into *E. coli* and expression was induced with IPTG (0.4 mM) for 3 h at 37 °C. The cells were then collected and lysed in binding buffer (10 mM Tris–HCl (pH 7.5), 100 mM NaCl, 10  $\mu$ g ml<sup>−1</sup> BSA, 10% glycerol, 0.05% NP-40). Lysates were sonicated, cleared by centrifugation (10,000g for 10 min) and incubated with ssTEL- (50  $\mu$ M) conjugated M280 beads (100  $\mu$ l) for 30 min at room temperature. The ssTEL and associated protein were captured by magnets, washed in binding buffer and eluted with 1 M NaCl for 10 min at 4 °C. The eluted protein was then diluted in binding buffer without salt to bring the final NaCl concentration down to 100 mM. *E. coli* single-stranded binding protein (SSB) was purchased from Promega.

**Gel-shift assay.** The 18-nucleotide telomeric ssDNA probe [(TTAGGG)<sub>3</sub>] was radiolabelled with  $\gamma$ -<sup>32</sup>P using T4 kinase and purified over a G25 column. The labelled ssDNA was incubated with purified RPA or POT1–TPP1 in binding buffer (10 mM Tris–HCl (pH 7.5), 100 mM NaCl, 10  $\mu$ g ml<sup>−1</sup> BSA, 10% glycerol, 0.05% NP-40) for 30 min at room temperature. The resulting protein–DNA complexes were separated by gel electrophoresis using 0.8% agarose at 140 V for 1.5 h and bands were visualized by autoradiography.

**DNA–protein binding assay using biotinylated ssDNA.** Biotinylated ssTEL [(TTAGGG)<sub>8</sub>] or ssMUT [(TTTGCG)<sub>8</sub>] were attached to streptavidin-coated magnetic beads in 10 mM Tris–HCl (pH 8.0), 100 mM NaCl at room temperature for 30 min. To analyse the bindings of purified RPA, POT1–TPP1 and POT1 to ssDNA, biotinylated ssDNA (1 pmol) was incubated with various amounts of purified protein in 500  $\mu$ l of binding buffer. To analyse the binding of RPA and Flag–POT1 to ssDNA in extracts, biotinylated ssDNA (10 pmol) and various amounts of extracts were added to 500  $\mu$ l of binding buffer. After incubation for 30 min, the protein–DNA complexes were retrieved with a magnet and washed three times with binding buffer. In the experiments using RPA or POT1 pre-coated ssDNA, biotinylated ssDNA (1 pmol) was first incubated with purified protein (3.8 pmol) for 30 min at room temperature. The ssDNA pre-coated with RPA or POT1 was retrieved with a magnet and subsequently mixed with increasing concentrations of WCE, nuclear extract or cytoplasmic extract for 30 min at room temperature. For nuclear extract inhibited by addition of TERRA or its derivatives ((UUAGGG)<sub>3</sub>, (CCCAUU)<sub>3</sub> and (UUGGCG)<sub>3</sub>), extracts were incubated with 1, 2, 5 or 10 pmol RNA for 30 min at 4 °C.

For hnRNPA1 binding, RPA-coated ssTEL or ssMUT (0.8 nM), or Flag–POT1-coated ssTEL (0.8 nM), were incubated with increasing concentrations of hnRNPA1 purified from *E. coli* (2.4, 4.8, 7.2 nM) and the proteins remaining on ssTEL were analysed by western blot. For TERRA inhibition, hnRNPA1 was pre-incubated with increasing concentrations of TERRA (2, 4, 10, 20 nM), or control RNA (UUGGCG)<sub>3</sub>. hnRNPA1 was then incubated with RPA coated ssTEL (0.8 nM). Similarly, to demonstrate that TERRA promotes the dissociation of hnRNPA1 from ssTEL, the ssTEL (0.8 nM) was pre-coated with hnRNPA1 (2.4 nM) and subsequently incubated with increasing concentrations of TERRA (2, 20, 200, 2,000 nM). To demonstrate that TERRA enhances POT1 binding, ssTEL (0.8 nM) was pre-coated with hnRNPA1 (2.4 nM) then incubated with both POT1 (2.4 nM) and increasing concentrations of TERRA (2, 20, 200 nM). In all reactions, the proteins remaining on DNA were analysed by western blot.

**Cell synchronization.** To follow the progression of cells from S to G2 (Fig. 4b, e), HeLa cells were synchronized with 2 mM thymidine for 16 h, washed three times with PBS and once with thymidine-free medium, and released into thymidine-free medium. To enrich HeLa cells in S phase of the cell cycle (Fig. 4a), cells were either collected after treatment for 16 h with 2 mM thymidine (early S), or collected 4 h after thymidine release (late S). To enrich cells in G1 and M phases, cells were either collected after treatment for 16 h with 0.1  $\mu$ g ml<sup>−1</sup> nocodazole (M), or collected 4 h after nocodazole release (G1).

**Extract preparation.** WCEs were either generated with the NETN buffer as described in the protein purification section, or with the binding buffer used in the DNA binding assays. Nuclear extract and cytoplasmic extract were generated as previously described<sup>3</sup>. To treat extracts with TERRA or its derivative RNA, RNA was added to WCE or nuclear extract in increasing concentrations (1, 2, 5, 10 pmol) and incubated for 30 min on ice.

**Capture of RPA-displacing activity from extracts.** To capture the RPA-displacing activity from extracts, RPA-coated ssTEL was incubated with nuclear extract for 30 min at room temperature. The beads were collected, washed three times in binding buffer, and eluted using the binding buffer with 1 M NaCl for 10 min on ice. The eluted material was collected, and diluted with the binding buffer without NaCl to reach a final NaCl concentration of 100 mM. The elution was incubated on ice for 1 h then added to RPA-coated ssDNA and incubated for 30 min at room temperature. For TERRA inhibition, either TERRA (UUAGGG)<sub>3</sub> or its derivative (CCCAUU)<sub>3</sub> were incubated with the eluted proteins before their addition to RPA-coated ssDNA. The proteins remaining bound to DNA were analysed by western blot.

**Identification of the RPA-displacing factors from extracts.** Biotinylated ssTEL or ssMUT (20 pmol) was attached to streptavidin-coated beads and coated with recombinant RPA. The RPA-coated ssTEL or ssMUT was incubated with 65  $\mu$ g of nuclear extract in 500  $\mu$ l of binding buffer. Beads with no DNA attached were used as a negative control. After 30 min of incubation, the beads were retrieved and washed three times with binding buffer containing 300 mM NaCl. The proteins associated with the RPA activity were eluted by binding buffer containing 600 mM NaCl for 10 min on ice. The eluted proteins from ssTEL, ssMUT and naked beads were separated by SDS–PAGE. After the gel was silver-stained, the two ~30-kDa bands specifically captured by RPA–ssTEL were excised and analysed by mass spectrometry.

**Immunofluorescence analysis.** HeLa cells were seeded onto coverslips and cultured overnight. The adhered cells were transfected with POT1 siRNA using oligofectamine (Invitrogen), or with hnRNPA1 siRNA using Lipofectamine RNAi Max (Invitrogen) and cultured for another 48 h. Synchronized cells were treated after 24 h with 2 mM thymidine for 16 h, washed and released, and processed at the indicated time points. Cells were extracted with 0.25% Triton, fixed in 3% paraformaldehyde and further permeabilized with 0.5% Triton. Cells were subsequently incubated with the primary antibodies (diluted in PBS containing 3% BSA and 0.05% Tween 20) for 1 h at 37 °C in a humidified chamber. After extensive washing with PBS, cells were incubated with secondary antibodies for 45 min at room temperature, and washed again with PBS. After incubation for 5 min with DAPI, cells were mounted on slides with Vectashield. Slides were analysed using a Nikon H600L fluorescence microscope.

**Combined immunofluorescence–RNA fluorescence in situ hybridization.** Cells were grown on coverslips and collected at different time points 17 h after release of single thymidine block. Cells were washed twice with cold PBS for 5 min and treated with cytochrome (100 mM NaCl, 300 mM sucrose, 3 mM MgCl<sub>2</sub>, 10 mM PIPES pH 7, 0.1% Triton X-100, 200 mM vanadyl ribonucleoside complex) for 7 min at 4 °C. Cells were rinsed briefly, fixed with 4% paraformaldehyde in PBS (USB 19943) for 10 min at room temperature. Cells were then washed three times with PBS for 5 min each and permeabilized with 0.5% NP40 in PBS for 10 min. Cells were washed twice with PBS for 5 min each and incubated with blocking solution (0.2% fish gelatin and 0.5% BSA) for 1 h. Cells were then incubated with human TRF2 antibody (clone 4A794 Upstate) at 1:2,000 and diluted in blocking solution for 2 h. After washing three times with PBST (PBS containing 0.1% Triton) for 10 min each, the cells were then incubated with secondary antibody Alexa 488 (Invitrogen A11001) at 1:2,000 dilution in blocking solution for 1 h. Cells were washed three times with PBST for 10 min each and were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature. Cells were rinsed briefly with PBS then incubated with hybridization mix (10 nM PNA–TAMRA–(CCCTAA) probe, 50% formamide, 2 $\times$  SSC, 2 mg ml<sup>−1</sup> BSA, 10% dextran sulphate, 10 mM vanadyl ribonucleoside complex) for 18 h in a humidified chamber at 39 °C. Cells were washed with 2 $\times$  SSC in 50% formamide three times at 39 °C for 5 min each, three times in 2 $\times$  SSC at 39 °C for 5 min each, and finally once in 2 $\times$  SSC at room temperature for 10 min. Coverslips were then mounted on glass microscope slides with Vectashield mounting medium containing DAPI (H-1200). For RNaseA treatment, coverslips were incubated with 200  $\mu$ g ml<sup>−1</sup> RNase A for

30 min at 37 °C before hybridization. Images were captured with an Endore cooled CCD (charge-coupled device) camera on a Nikon eclipse 80i microscope and the images processed with NIS-Element BR 3.10 software.

**Chromatin immunoprecipitation.** RPA chromatin immunoprecipitation and the analysis of telomere association were performed as previously described<sup>7</sup>. Cells were transfected twice with hnRNPA1 siRNA (hnRNPA1-1) and synchronized with thymidine for 15 h. The two RPA2 antibodies used were from Abcam and Thermo.

**Antibodies and siRNA.** The RPA pS33 antibody was from Bethyl. The monoclonal antibody to RPA2 was from Neomarkers. The anti-FLAG M2 antibody was from Sigma. The Chk1 antibody and Cyclin A antibody were from Santa Cruz, and

the phospho-Chk1 Ser345 antibody was from Cell Signaling. The TRF2 antibody was from Bethyl. The phospho-H3 Ser10 antibody was from Millipore. The H4 antibody was from Active Motif. The hnRNPA1 antibody was from Cell Signaling. The POT1 siRNA used in Fig. 4e and Supplementary Fig. 10 was the SMARTPOOL from Dharmacon. The hnRNPA1 siRNAs used in Fig. 4c, d and Supplementary Figs 8 and 9 were CAACUUCGGUC-GUGGAGGA and UCCACGACCACCACCAAAG.

30. Henriksen, L. A. & Wold, M. S. Replication protein A mutants lacking phosphorylation sites for p34cdc2 kinase support DNA replication. *J. Biol. Chem.* **269**, 24203–24208 (1994).



# CAREERS

**TURNING POINT** Fellowship opens research paths for computational biologist **p.539**

**BRIEF** NASA launches website to draw women's interest **p.539**

**NATUREJOBS** For the latest career listings and advice [www.naturejobs.com](http://www.naturejobs.com)

IMAGES.COM/CORBIS



## CANCER RESEARCH

# Promise of protection

*Cancer vaccines have long shown lots of potential but few results. Signs of success now suggest opportunities.*

BY KELLY RAE CHI

Last April, after years of false starts and frustration, the first-ever therapeutic cancer vaccine gained approval from the US Food and Drug Administration (FDA). Provenge (sipuleucel-T), made by Dendreon in Seattle, Washington, has been demonstrated to extend the survival of patients with late-stage

prostate cancer by several months, and its success has shown the way for other cancer vaccines, says James Gulley, a medical oncologist at the National Cancer Institute in Bethesda, Maryland. "We knew there were regulatory challenges that the FDA needed to look at to get this approved," he says. "So it set the milestones for future paths for our therapeutic vaccines."

Cancer vaccines have had a boost in the past

three years, thanks in part to Provenge's success as well as to a growing number of promising therapies in late-stage clinical trials, says Gulley. Pharmaceutical companies have approached him about developing therapeutic-vaccine commercialization programmes, and small biotechnology firms have shown interest in combining cancer-drug projects with vaccine development.

Hundreds of clinical trials of cancer vaccines are currently under way in the United States, and Hyam Levitsky, an oncologist at the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University in Baltimore, Maryland, sees more promise now than at any time during his 20 years in the field. People thinking about going into cancer-vaccine work can be assured that "there is going to be a future in this", he says.

Levitsky and his colleagues are optimistic that cancer vaccines will find their way into the standard of care not only because of the success of Provenge but also because of the wealth of trials going on. All this signals research opportunities in the near future. Some of the available jobs, from discovery to manufacturing and testing, are the same as those in the field of conventional vaccines. But cancer vaccines — immunotherapies that trigger or enhance defences against tumours or cancer-causing pathogens — are particularly challenging to develop, and most are still in the experimental stages. They work through a variety of mechanisms, making it difficult for scientists to develop standards to test and approve them. For the best chance of advancing in the field, researchers must have a solid grounding in immunology and the ability to understand how cancers interact with the immune system.

## BENCH TO BEDSIDE

Cancer vaccines fall into two major categories: prophylactic vaccines, which are given to healthy people to prevent cancer developing, and therapeutic vaccines, which are intended to treat an existing cancer by strengthening the body's natural defences. Therapeutic vaccines are usually made from cancer cells, or from viruses or molecules present in tumours of some or many types of cancer, whereas preventive vaccines typically target viruses that cause cancer. Most vaccines of both types are developed using a one-size-fits-all approach, such as targeting molecules that are present in the cancers of many patients, but some, including Provenge, are tailored to an individual ►

► patient.

Compared with infectious diseases, cancer poses some extra challenges for vaccine-makers — enough to give pause even to optimistic scientists and clinicians. Investigators must ensure that their vaccines target tumour-specific cells without triggering a more general immune response against normal cells. Whereas traditional

prophylactic vaccines prevent infections in healthy people, therapeutic cancer vaccines must be given to patients whose immune systems have been weakened by conventional cancer treatments such as chemotherapy. Vaccines can take time to work, so they offer small comfort for those with only a few months to live.

And creating personalized therapies entails further scientific and commercial difficulties: it often requires harvesting the patient's own cancer cells or developing companion diagnostics to identify subsets of patients who might benefit from the therapy. "Doing large trials and getting drugs approved and funded when they're individualized is very, very difficult," says John Sampson, a neuro-oncologist at Duke University Medical Center in Durham, North Carolina.

Anyone wishing to develop cancer vaccines must have a understand tumour immunology, because much of the basic science focuses on understanding the role of the immune system and inflammation in the development and progression of cancer. Only a few institutions offer doctoral programmes specifically in cancer immunology; they include the University of Lausanne in Switzerland and the State University of New York at Buffalo. The majority of cancer immunologists come from either immunology or cancer-biology programmes.

Newcomers to the field should keep in mind that cancer-drug development is risky: it can take a decade or longer for a drug to get from the lab to the clinic. "Your chances of getting something that works are extraordinarily small," says Sampson. Provenge needed 20 years and three phase III clinical trials to reach FDA approval.

Most big pharmaceutical companies are still watching from the sidelines, choosing not to fund or hire for early-stage development, says Karl-Josef Kallen, chief scientific officer of CureVac, a biotechnology firm based in Tübingen, Germany. Vincent Tuohy, an immunologist at the Cleveland Clinic Lerner Research Institute in Ohio, has struggled to raise funds for a phase I clinical trial of a prophylactic vaccine aimed at women with a high risk of

developing breast cancer. "I'm having a rough time trying to convince the pharmaceutical industry, whose business model is based predominantly on treatment and diagnostics, to enter the preventive space for these cancers," says Tuohy, who has also been turned away by government funding agencies and non-profit organizations.

One exception is GlaxoSmithKline (GSK), whose immunotherapeutics division is looking for clinical development managers for early- and late-stage cancer-vaccine development, medical affairs and project management at its biologics research hub in Rixensart, Belgium. The ideal candidates have experience in immunotherapeutics or general oncology, or specialize in specific types of cancer, says Roya Paganini, senior manager in charge of talent acquisition at GSK Biologicals in Rixensart.

Smaller, discovery-focused biotechnology companies may have more opportunities. CureVac, for example, is looking for PhD-level research scientists to design vaccines and plan experiments to test them. They also need to be able to tweak the therapy's dose and dosing schedule in combination with existing and new therapies, to demonstrate its efficacy.

#### TRIAL RUNS

Most training in designing and running clinical trials for cancer-drug development is simply not that helpful for vaccine development. Unlike traditional cancer therapies, for example, vaccines do not have a 'maximum tolerated dose'. In fact, the appropriate dosing, scheduling of treatment and end points for vaccines all differ vastly from those for traditional drugs.

That makes immunology training, whether formal or on-the-job, essential for success. Principal investigators who are involved in patient care usually have medical doctorates, board certification in internal medicine or paediatrics, and experience in an oncology fellowship lasting several years. Immunology is not a distinct subspecialty of oncology, so researchers should decide early on whether clinical research centred on cancer immunology is a path they wish to pursue. During an oncology fellowship, scientists with an interest in cancer vaccines should seek a mentor with experience as a faculty investigator, says Levitsky, adding that fellows might expect to design and help to analyse the results of one or two trials. By the time Gulley had finished his fellowship at the National Institutes of Health in Bethesda, he had nearly completed one clinical trial and had ideas for several more. Fellows training with Levitsky also sit in on first-year medical-school immunology courses and attend specialized seminars and journal clubs to stay up-to-date in the field.

Trials need skilled clinical-research coordinators, who often work with or under the principal investigator to speed recruitment

or reduce costs in clinical trials. "I think we need to find people who are much better at figuring out ways to do clinical trials well but efficiently," says Sampson, who has graduated from a 'master of health sciences in clinical research' programme at Duke. The university also offers a non-degree option for clinicians, nurses and scientists.

#### MAKING VACCINES

In principle, making cancer vaccines is not that different from making monoclonal antibodies or other biologics, so one need not have a cancer-research background to get into manufacturing, quality control or quality assurance, says George Mitra, director of the Biopharmaceutical Development Program at the National Cancer Institute's facility in Frederick, Maryland. However, individualized vaccines can be more like a specialized research project, in which clinicians collect cells from a patient with cancer and use them to make a vaccine for that patient. Making such vaccines requires more skilled labour and experience than making generalized vaccines, says Mitra. Creating Provenge, for example, involves a process called leukapheresis, in which antigen-presenting cells are harvested from a patient's blood and purified. The cells are then sent to a Dendreon facility, where researchers cultivate them with a proprietary manufactured protein commonly found on prostate-cancer cells. The resulting vaccine is returned to the patient's physician. It must be administered in three different treatments, each of which requires the same manufacturing process.



**"There is going to be a future in this."**

Hyam Levitsky

Although the odds are long, if an individualized vaccine becomes commercially feasible, industrial manufacture would require PhD- and master's-level scientists with a background in cellular processing, who are trained in blood apheresis, bone-marrow transplantation or graft engineering,

as opposed to more traditional drug manufacturing areas, says Levitsky.

If the field continues to grow, interested scientists will have options. But the field's biggest draw may be the potential to help desperate patients. "To me, it's immensely rewarding to see that we can make a difference in patients' outcomes and survival, without causing significant side effects," says Gulley. "That's what's going to keep me working in this field and make me try for bigger improvement in outcomes for these patients." ■ **SEE OUTLOOK P.450**

**Kelly Rae Chi** is a freelance journalist based in Cary, North Carolina.

# TURNING POINT

## Richard Green

*Richard Green, a computational biologist at the University of California, Santa Cruz, was one of 118 young researchers to win a US\$50,000 two-year research fellowship from the Alfred P. Sloan Foundation, a philanthropic institution based in New York City. Green tells Nature how the fellowship will help to distinguish his work from that of his mentors.*

### Why did you choose computational biology over bench work?

I found that the cycle of determining whether an experimental idea is good or bad was much faster with a computer than a pipette — which is a big deal for someone like me, with many varied interests I hope to explore.

### What is the market like for computational biologists?

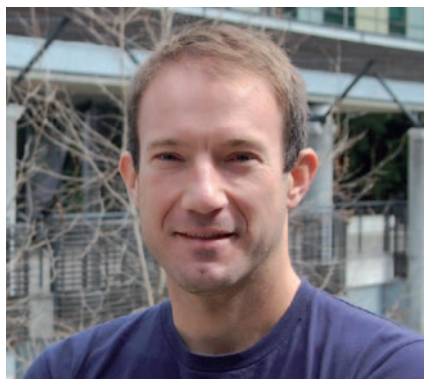
There is a huge demand, but not that many people who marry a deep understanding of molecular biology with the ability to think in terms of algorithms. It's really a seller's market for people with these skills. As a result, I'm in the enviable position of getting to pick and choose what I want to work on. It's important to note that there is not some new discipline called 'computational biology'; it's really a third entity that combines biology and computational ability. I suggest that young scientists become specialists in both disciplines.

### Has the prolific generation of genomics data altered career expectations?

The bar has been raised: the sequencing and basic analysis of a genome is no longer an automatic paper in *Science* or *Nature*. But that's a good thing. We have to be more creative; for example, species' genomes can be compared to gain an evolutionary perspective of the transition to multicellularity.

### How did you navigate your postdoc?

If grad school is where a young scientist's ship is assembled, then the postdoc is the launching pad where you take off — or not. I had a great set of projects as a graduate student in Steven Brenner's lab at the University of California, Berkeley. One of my last projects was investigating how alternative gene splicing evolves in flies. As a postdoc, I wanted to ask the next obvious question: how quickly did alternative splicing evolve between chimps and humans? I was fortunate that my adviser introduced me to Svante Pääbo, director of evolutionary genetics at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.



Having your boss put you on these top researchers' radar screens is incredibly helpful.

### How did you get involved with the Neanderthal genome sequencing project?

Blind luck. Shortly after I arrived, Svante was experimenting with a new shotgun sequencing method before applying it to the precious few samples of Neanderthal DNA extracted from bone remnants. He wanted someone to align the sequences of ancient cave-bear DNA derived using this method to other genome sequences in the database — something I could easily do, so I volunteered to impress my boss. Those alignments proved that the method worked. We then applied it to some of our best Neanderthal extracts, and Svante encouraged me to work on this. The Neanderthal project was such an incredible opportunity that I left the alternative splicing to the side.

### Did that cause problems with funders?

A little. I dutifully told my funders about my change in focus — and I had to reapply under a different programme to avoid getting my funding pulled. It was unfortunate that the funders didn't immediately see that this was a once-in-a-lifetime project and that I was perfectly positioned to do it.

### Is your work riskier than most scientists'?

Maybe. One of the hallmarks of my research is that I do many things. This traditionally has been viewed as a weakness. If you have many areas of expertise, they can erode one another in people's perceptions. I want to use my expertise to move different fields forward; for example, sequencing the alligator genome will offer new insight into developmental as well as evolutionary biology. ■

INTERVIEW BY VIRGINIA GEWIN

## LEGISLATION

### Gender equality bid

For the third time since 2008, US legislators have introduced a bill in the House of Representatives aimed at tackling gender inequalities in research. Eddie Bernice Johnson (Democrat, Texas), the sponsor of the bill, argues that the paucity of women in fields such as physics constitutes a competitive disadvantage for the United States. The bill would compel all US funding agencies to extend grants or fund temporary workers when researchers need time off to care for their families. The first try at the legislation died in committee in 2008; the second, last year, hitched it to another bill from which it was stripped before passage. Progress depends on the House Committee on Science, Space, and Technology and its chairman, Ralph Hall (Republican, Texas), who will decide whether the bill will be considered.

## BIOMEDICAL FUNDING

### Disheartening cuts

The effects of a US budget crunch could drive young scientists into other jobs, warns Francis Collins, director of the National Institutes of Health (NIH). At a Washington DC panel on 15 March, Collins noted that NIH grant success rates have slid from 25–30% to 20% and below in the past 4 years, owing to cuts. "Are these people going to keep banging their heads against the wall? Or are they going to find some other way to make a living?" said Collins at the talk, which was presented by Research!America, an advocacy group in Alexandria, Virginia. Few legislators understand the importance of investing in research, said panelist and former US representative Mike Castle. "Agencies need to get their story out," he said.

## NASA

### Agency woos women

NASA has built a website to boost women's and girls' familiarity with the agency, raise their interest in working there and increase recognition of female contributions to aeronautics. The site, <http://women.nasa.gov>, was launched on 16 March and gives glimpses into the work lives and accomplishments of female astronauts, researchers and engineers through videos and essays. A careers page provides links to job openings, including graduate and postdoc fellowships. A NASA spokeswoman says that the agency is trying to change the perception that it is male-dominated.



# RENEWAL

*Flesh of my flesh.*

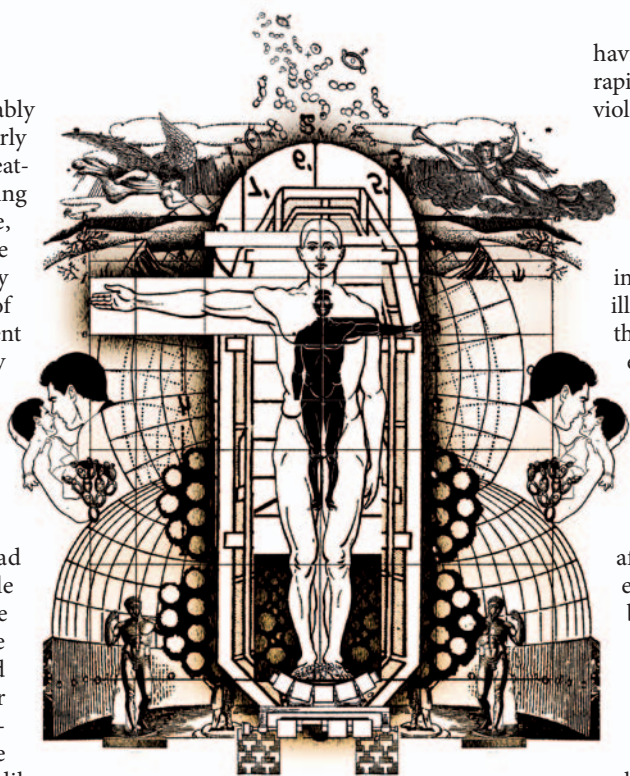
BY PETER ROBERTS

The waiting room was comfortably cool. The whole place had clearly been engineered to be non-threatening, even calming, while still looking professional: the walls were a pale, matte-textured powder blue, and the furniture and accessories were tastefully nondescript in various darker shades of blue. There was an air-conditioning vent on the wall, not quite directly above my head. The outflow gently touched my face, almost unnoticed, much as apprehension gently touched my consciousness.

My anticipation prior to the procedure was probably more intense than that of the typical patient. I had more to pay attention to. Most people focus on the anticipated benefits of the procedure, with their hopeful attitude perhaps slightly tempered by a mild anxiety about possible side effects or bad outcomes — a wholly unjustified anxiety if you ask me, as no one has ever experienced any adverse bodily consequences, and virtually everyone has emerged healthier (often significantly so). Those who undergo treatment when they're still young and healthy may not receive much benefit, but they are not harmed. I should know — I'm part of the team that developed the procedure.

And because of my role in the development team, I was, unlike most of the others patiently awaiting their turn in the bioculture chamber, familiar with the reports of bad psychological reactions to the procedure. These traumatic incidents, the only reported negative outcomes, are relatively rare, but can be quite severe, even somewhat debilitating, at least for a time. So, as the eldest member of the development team, and therefore the first to undergo the procedure, I was ready to observe and report on my emotional reaction. I thought I had prepared myself adequately. I was wrong.

Of course, I knew that the procedure would be complicated: age-asymmetric budding does not come as naturally to humans as it does to yeasts. And we had inverted the process as well, further complicating matters. Virtually every aspect of human physiology had to be modified, manipulated or suspended temporarily during the course of treatment. But the end result was unquestionably worth all the complications. Yeast



budding leaves a parent cell full of the miscoded and broken genes, somatic impurities and contamination, and telomere shortening that ageing induces (or that produce ageing), while generating as a result a youthful, near-perfect offspring. We had altered the process, so that the human bud would contain all the impurities, miscodings and so forth, leaving a purified and rejuvenated parent. This seemed to us to be a major step, perhaps even the final one, on the road to human immortality. Surely these results, these benefits, more than compensated for any psychological stress.

And yet, shockingly, some of those who experienced the emotional side effects refused to repeat the treatment. They had become as adamantly opposed to the whole enterprise as those religious fanatics (dwindling in absolute numbers, but not, it seemed, in influence) who objected to what they saw as 'playing God'.

I won't bother you with a detailed description of the preparation, or of the actual procedure, except to say that the discomforts were mild, and the indignities nothing out of the ordinary for this sort of situation. What was out of the ordinary was the nearly indescribable sensation of

having something 'other' growing, rather rapidly, out of my body — a sort of intimate violation, but without any sense of violence.

My first thought was, "This must be what having a baby is like!" But of course I recognized that wasn't quite accurate. It was more like having an identical twin, a conjoined twin — an incomplete, profoundly deformed, deathly ill twin, but a twin no less for all that. And then, shortly after the twin came into existence, he had to be surgically removed and destroyed. After undergoing the experience, I was not surprised that some people found it deeply disturbing; rather, I was amazed that so few people were adversely affected.

I did what I suppose most people do afterwards: I intellectualized the experience. I went through the process step by step, reminding myself that the twin was genetically identical to me (aside from the miscodings and so on), and derived directly from my own body, like a tumour or a wart. My bud had no mental function to speak of, and hence didn't suffer, or experience anything whatsoever. Excising and disposing of my twin was not very different from what would have been done with any other morbid lesion.

But that's not how it felt.

My follow-up report did recommend one change to the aftercare protocol. I suggested that, if they so chose, patients should be permitted to stay with their bud, to nurse and nurture it until its natural demise. This, I hoped, would be less of a shock than having what had been an intimate part of oneself suddenly snatched away. At any rate, the bud's survival time was likely to be quite brief. It might even be healthy for people to have the experience of taking care of — not exactly a child, but probably the closest thing to a child most of us will ever encounter. Also, morbid though it may seem, this altered protocol would provide people with perhaps their only experience of serious illness, death and mourning.

I could muster a variety of perfectly rational arguments in favour of this protocol change, but in the end, for me, the most powerful argument came down to this: it just seemed like the right thing to do. ■

*Peter Roberts is a prolific poet: his poems and stories have been published in a wide variety of magazines and journals.*

➤ **NATURE.COM**  
Follow Futures on  
Facebook at:  
[go.nature.com/mtoodm](https://www.facebook.com/mtoodm)

Inclusive fitness theory and eusociality

ARISING FROM M. A. Nowak, C. E. Tarnita & E. O. Wilson *Nature* **466**, 1057–1062 (2010)

Nowak *et al.*<sup>1</sup> argue that inclusive fitness theory has been of little value in explaining the natural world, and that it has led to negligible progress in explaining the evolution of eusociality. However, we believe that their arguments are based upon a misunderstanding of evolutionary theory and a misrepresentation of the empirical literature. We will focus our comments on three general issues.

First, Nowak *et al.*<sup>1</sup> are incorrect to suggest a sharp distinction between inclusive fitness theory and “standard natural selection theory”. Natural selection explains the appearance of design in the living world, and inclusive fitness theory explains what this design is for. Specifically, natural selection leads organisms to become adapted as if to maximize their inclusive fitness<sup>2–4</sup>. Inclusive fitness theory is based upon population genetics, and is used to make falsifiable predictions about how natural selection shapes phenotypes, and so it is not surprising that it generates identical predictions to those obtained using other methods<sup>2,5–7</sup>.

Second, Nowak *et al.*<sup>1</sup> are incorrect to state that inclusive fitness requires a number of “stringent assumptions” such as pairwise interactions, weak selection, linearity, additivity and special population structures. Hamilton’s original formulations did not make all these assumptions, and generalizations have shown that none of them is required<sup>3,5,6,8</sup>. Inclusive fitness is as general as the genetical theory of natural selection itself. It simply partitions natural selection into its direct and indirect components.

Nowak *et al.*<sup>1</sup> appear to have confused the completely general theory of inclusive fitness with models of specific cases. Yes, researchers often make limiting assumptions for reasons of analytical tractability when considering specific scenarios<sup>5,7</sup>, as with any modelling approach. For example, Nowak *et al.*<sup>1</sup> assume a specific form of genetic control, where dispersal and helping are determined by the same single locus, that mating is monogamous, and so on. However, the inclusive fitness approach has facilitated, not hindered, empirical testing of evolutionary theory<sup>9–11</sup>. Indeed, an advantage of inclusive fitness theory is that it readily generates testable predictions in situations where the precise genetic architecture of a phenotypic trait is unknown.

Third, we dispute the claim of Nowak *et al.*<sup>1</sup> that inclusive fitness theory “does not provide any additional biological insight”, delivering only “hypothetical explanations”, leading only to routine measurements and “correlative studies”, and that the theory has “evolved into an abstract enterprise largely on its own”, with a failure to consider multiple competing hypotheses. We cannot explain these claims, which seem to overlook the extensive empirical literature that has accumulated over the past 40 years in the fields of behavioural and evolutionary ecology<sup>9–11</sup> (Table 1). Of course, studies must consider the direct consequences of behaviours, as well as consequences for relatives, but no one claims otherwise, and this does not change the fact that relatedness (and lots of other variables) has been shown to be important in all of the above areas.

We do not have space to detail all the advances that have been made in the areas described in Table 1. However, a challenge to the claims of Nowak *et al.*<sup>1</sup> is demonstrated with a single example, that of sex allocation (the ratio of investment into males versus females). We choose sex allocation because: (1) Nowak *et al.*<sup>1</sup> argue that inclusive fitness theory has provided only “hypothetical explanations” in this field; (2) it is an easily quantified social trait, which inclusive fitness theory predicts can be influenced by interactions between relatives; and (3) the study of sex allocation has been central to evolutionary work on the eusocial insects. In contrast to the claims of Nowak *et al.*<sup>1</sup>,

recent reviews of sex allocation show that the theory explains why sex allocation varies with female density, inbreeding rate, dispersal rate, brood size, order of oviposition, sib-mating, asymmetrical larval competition, mortality rate, the presence of helpers, resource availability and nest density in organisms such as protozoan parasites, nematodes, insects, spiders, mites, reptiles, birds, mammals and plants<sup>5,12,13</sup>.

The quantitative success of this research is demonstrated by the percentage of the variance explained in the data. Inclusive fitness theory has explained up to 96% of the sex ratio variance in across-species studies and 66% in within-species studies<sup>13</sup>. The average for all evolutionary and ecological studies is 5.4%. As well as explaining adaptive variation in behaviour, inclusive fitness theory has even elucidated when and why individuals make mistakes (maladaptation), in response to factors such as mechanistic constraints<sup>13</sup>. It is not clear how Nowak *et al.*<sup>1</sup> can characterize such quantifiable success as “meagre”. Their conclusions are based upon a discussion in the Supplementary Information of just three papers (by authors who disagree with the interpretations of Nowak *et al.*<sup>1</sup>), out of an empirical literature of thousands of research articles. This would seem to indicate a failure to engage seriously with the body of work that they recommend we abandon.

The same points can be made with regard to the evolution of the eusocial insects, which Nowak *et al.*<sup>1</sup> suggest cannot be explained by inclusive fitness theory. It was already known that haplodiploidy itself may have only a relatively minor bearing on the origin of eusociality, and so Nowak *et al.*<sup>1</sup> have added nothing new here. Inclusive fitness theory has explained why eusociality has evolved only in monogamous lineages, and why it is correlated with certain ecological conditions, such as extended parental care and defence of a shared resource<sup>14,15</sup>. Furthermore, inclusive fitness theory has made very successful predictions about behaviour in eusocial insects, explaining a wide range of phenomena (Table 2).

Ultimately, any body of biological theory must be judged on its ability to make novel predictions and explain biological phenomena; we believe that Nowak *et al.*<sup>1</sup> do neither. The only prediction made by their model (that offspring are favoured to help their monogamously

Table 1 | Inclusive fitness theory has been important in understanding a range of behavioural phenomena

| Research area             | Correlational? | Experimental? | Theory–data interplay |
|---------------------------|----------------|---------------|-----------------------|
| Sex allocation            | Yes            | Yes           | Yes                   |
| Policing                  | Yes            | Yes           | Yes                   |
| Conflict resolution       | Yes            | Yes           | Yes                   |
| Cooperation               | Yes            | Yes           | Yes                   |
| Altruism                  | Yes            | Yes           | Yes                   |
| Spite                     | Yes            | Yes           | Yes                   |
| Kin discrimination        | Yes            | Yes           | Yes                   |
| Parasite virulence        | Yes            | Yes           | Yes                   |
| Parent–offspring conflict | Yes            | Yes           | Yes                   |
| Sibling conflict          | Yes            | Yes           | Yes                   |
| Selfish genetic elements  | Yes            | Yes           | Yes                   |
| Cannibalism               | Yes            | Yes           | Yes                   |
| Dispersal                 | Yes            | Yes           | Yes                   |
| Alarm calls               | Yes            | Yes           | Yes                   |
| Eusociality               | Yes            | Yes           | Yes                   |
| Genomic imprinting        | Yes            | Yes           | Yes                   |

Data are taken from refs 9–11. Correlational studies test predictions using natural variation in key variables, whereas experimental studies involve their experimental manipulation. Interplay between theory and data means that theory has informed empirical study, and vice versa. Inclusive fitness is not the only way to model evolution, but it has already proven to be an immensely productive and useful approach for studying eusociality and other social behaviours.

# BRIEF COMMUNICATIONS ARISING

**Table 2 | Areas in which inclusive fitness theory has made successful predictions about behaviour in eusocial insects**

| Trait examined                                      | Explanatory variables   | Correlational studies? | Experimental studies? | Interplay between theory and data? |
|---|---|------------------------|-----------------------|------------------------------------|
| Altruistic helping                                  | Haplodiploidy versus diploidy   | Yes                    | No                    | Yes                                |
| Worker egg laying                                   | Worker policing   | Yes                    | Yes                   | Yes                                |
| Policing  | Relatedness   | Yes                    | Yes                   | Yes                                |
| Level of cooperation                                | Costs, benefits and relatedness   | Yes                    | Yes                   | Yes                                |
| Intensity of work                                   | Need for work and probability of becoming queen   | Yes                    | Yes                   | Yes                                |
| Sex allocation                                      | Relatedness asymmetries due to variation in queen survival, queen number and mating frequency | Yes                    | Yes                   | Yes                                |
| Sex allocation                                      | Resource availability   | Yes                    | Yes                   | Yes                                |
| Sex allocation                                      | Competition for mates between related males   | Yes                    | Yes                   | Yes                                |
| Number of individuals trying to become reproductive | Presence of old queens  | Yes                    | Yes                   | Yes                                |
| Workers killing queens                              | Presence of workers, reproductives or other queens  | Yes                    | No                    | No                                 |
| Exclusion of non-kin                                | Colony membership   | Yes                    | Yes                   | Yes                                |

Data are taken from refs 12–16.

mated mother if this provides a sufficient benefit) merely confirms, in a less general way, Hamilton's original point: if the fitness benefits are great enough, then altruism is favoured between relatives.

Patrick Abbot<sup>1</sup>, Jun Abe<sup>2</sup>, John Alcock<sup>3</sup>, Samuel Alizon<sup>4</sup>, Joao A. C. Alpedrinha<sup>5</sup>, Malte Andersson<sup>6</sup>, Jean-Baptiste Andre<sup>7</sup>, Minus van Baalen<sup>7</sup>, Francois Balloux<sup>8</sup>, Sigal Balshine<sup>9</sup>, Nick Barton<sup>10</sup>, Leo W. Beukeboom<sup>11</sup>, Jay M. Biernaskie<sup>5</sup>, Trine Bilde<sup>12</sup>, Gerald Borgia<sup>13</sup>, Michael Breed<sup>14</sup>, Sam Brown<sup>5</sup>, Redouan Bshary<sup>15</sup>, Angus Buckling<sup>5</sup>, Nancy T. Burley<sup>16</sup>, Max N. Burton-Chellaw<sup>5</sup>, Michael A. Cant<sup>17</sup>, Michel Chapuisat<sup>18</sup>, Eric L. Charnov<sup>19</sup>, Tim Clutton-Brock<sup>20</sup>, Andrew Cockburn<sup>21</sup>, Blaine J. Cole<sup>22</sup>, Nick Colegrave<sup>23</sup>, Leda Cosmides<sup>24</sup>, Iain D. Couzin<sup>25</sup>, Jerry A. Coyne<sup>26</sup>, Scott Creel<sup>27</sup>, Bernard Crespi<sup>28</sup>, Robert L. Curry<sup>29</sup>, Sasha R. X. Dall<sup>17</sup>, Troy Day<sup>30</sup>, Janis L. Dickinson<sup>31</sup>, Lee Alan Dugatkin<sup>32</sup>, Claire El Mouden<sup>5</sup>, Stephen T. Emlen<sup>33</sup>, Jay Evans<sup>34</sup>, Regis Ferriere<sup>35</sup>, Jeremy Field<sup>36</sup>, Susanne Foitzik<sup>37</sup>, Kevin Foster<sup>5</sup>, William A. Foster<sup>20</sup>, Charles W. Fox<sup>38</sup>, Juergen Gadau<sup>39</sup>, Sylvain Gandon<sup>40</sup>, Andy Gardner<sup>5</sup>, Michael G. Gardner<sup>41</sup>, Thomas Getty<sup>42</sup>, Michael A. D. Goodisman<sup>43</sup>, Alan Grafen<sup>5</sup>, Rick Grosberg<sup>44</sup>, Christina M. Grozinger<sup>45</sup>, Pierre-Henri Gouyon<sup>46</sup>, Darryl Gwynne<sup>47</sup>, Paul H. Harvey<sup>5</sup>, Ben J. Hatchwell<sup>48</sup>, Jürgen Heinze<sup>49</sup>, Heikki Helanterä<sup>50</sup>, Ken R. Helms<sup>51</sup>, Kim Hill<sup>52</sup>, Natalie Jiricny<sup>5</sup>, Rufus A. Johnstone<sup>20</sup>, Alex Kacelnik<sup>5</sup>, E. Toby Kiers<sup>53</sup>, Hanna Kokko<sup>21</sup>, Jan Komdeur<sup>54</sup>, Judith Korb<sup>55</sup>, Daniel Kronauer<sup>56</sup>, Rolf Kümmerli<sup>57</sup>, Laurent Lehmann<sup>15</sup>, Timothy A. Linksvayer<sup>58</sup>, Sébastien Lion<sup>59</sup>, Bruce Lyon<sup>60</sup>, James A. R. Marshall<sup>61</sup>, Richard McElreath<sup>62</sup>, Yannis Michalakis<sup>4</sup>, Richard E. Michod<sup>63</sup>, Douglas Mock<sup>64</sup>, Thibaud Monnin<sup>7</sup>, Robert Montgomerie<sup>65</sup>, Allen J. Moore<sup>17</sup>, Ulrich G. Mueller<sup>66</sup>, Ronald Noë<sup>67</sup>, Samir Okasha<sup>68</sup>, Pekka Pamilo<sup>69</sup>, Geoff A. Parker<sup>70</sup>, Jes S. Pedersen<sup>58</sup>, Ido Pen<sup>71</sup>, David Pfennig<sup>72</sup>, David C. Queller<sup>73</sup>, Daniel J. Rankin<sup>74</sup>, Sarah E. Reece<sup>23</sup>, Hudson K. Reeve<sup>33</sup>, Max Reuter<sup>75</sup>, Gilbert Roberts<sup>76</sup>, Simon K. A. Robson<sup>77</sup>, Denis Roze<sup>78</sup>, Francois Rousset<sup>79</sup>, Olav Rueppell<sup>80</sup>, Joel L. Sachs<sup>81</sup>, Lorenzo Santorelli<sup>5</sup>, Paul Schmid-Hempel<sup>82</sup>, Michael P. Schwarz<sup>41</sup>, Tom Scott-Phillips<sup>83</sup>, Janet Shellmann-Sherman<sup>33</sup>, Paul W. Sherman<sup>33</sup>, David M. Shuker<sup>84</sup>, Jeff Smith<sup>73</sup>, Joseph C. Spagna<sup>85</sup>, Beverly Strassmann<sup>86</sup>, Andrew V. Suarez<sup>87</sup>, Liselotte Sundström<sup>50</sup>, Michael Taborsky<sup>88</sup>, Peter Taylor<sup>30</sup>, Graham Thompson<sup>89</sup>, John Tooby<sup>90</sup>, Neil D. Tsutsui<sup>91</sup>, Kazuki Tsuji<sup>92</sup>, Stefano Turillazzi<sup>93</sup>, Francisco Úbeda<sup>94</sup>, Edward L. Vargo<sup>95</sup>, Bernard Voelkl<sup>96</sup>, Tom Wenseleers<sup>97</sup>, Stuart A. West<sup>5</sup>, Mary Jane West-Eberhard<sup>98</sup>, David F. Westneat<sup>99</sup>, Diane C. Wiernasz<sup>22</sup>, Geoff Wild<sup>100</sup>, Richard Wrangham<sup>101</sup>, Andrew J. Young<sup>17</sup>, David W. Zeh<sup>102</sup>, Jeanne A. Zeh<sup>102</sup> & Andrew Zink<sup>103</sup>

<sup>1</sup>Vanderbilt University, Nashville, Tennessee 37235, USA.

<sup>2</sup>Laboratory of Applied Entomology, Faculty of Agriculture, Shizuoka University, Sizuoka 422-8529, Japan.

<sup>3</sup>School of Life Sciences, PO Box 874501, Arizona State University, Tempe, Arizona 85287-4501, USA.

<sup>4</sup>UMR CNRS-IRD 2724, Genetics and Evolution of Infectious Diseases, IRD, 911 Avenue Agropolis, B.P. 64501, 34394 Montpellier Cedex 5, France.

<sup>5</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK.

e-mail: Stuart.West@zoo.ox.ac.uk

<sup>6</sup>Department of Zoology, University of Gothenburg, SE 405 30 Gothenburg, Sweden.

<sup>7</sup>CNRS, Université Pierre et Marie Curie, Ecole Normale Supérieure, UMR 7625, Ecologie and Evolution, 75005 Paris, France.

<sup>8</sup>MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK.

<sup>9</sup>Department of Psychology, Neuroscience and Behaviour, McMaster University, 1280 Main St West, Hamilton, Ontario L8S 4K1, Canada.

<sup>10</sup>IST Austria, Am Campus 1, Klosterneuburg 3400, Austria.

<sup>11</sup>Evolutionary Genetics, Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 14, NL-9750 AA Haren, The Netherlands.

<sup>12</sup>Aarhus University, Department of Biological Sciences, Ny Munkegade 1540, 8000 Aarhus C, Denmark.

<sup>13</sup>Department of Biology, University of Maryland, College Park, Maryland 20742-4415, USA.

<sup>14</sup>Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309-0334, USA.

<sup>15</sup>Faculté des sciences, Rue Emile-Argand 11, Case postale 158, 2000 Neuchâtel, Switzerland.

<sup>16</sup>Department of Ecology and Evolutionary Biology, University of California, 321 Steinhaus Hall, Irvine, California 92697-2525, USA.

<sup>17</sup>Centre for Ecology and Conservation, University of Exeter, Cornwall, Tremough, Penryn TR10 9EZ, UK.

<sup>18</sup>Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.

<sup>19</sup>Department of Biology, 167 Castetter Hall, MSC03 2020, 1 University of New Mexico, Albuquerque, New Mexico 87131-0000, USA.

<sup>20</sup>Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK.

<sup>21</sup>Evolution, Ecology and Genetics, Research School of Biology, Australian National University, Canberra, ACT 0200, Australia.

<sup>22</sup>Department of Biology and Biochemistry, University of Houston, Houston, Texas 77204-5001, USA.

<sup>23</sup>Institutes of Evolution, Immunology and Infection Research, School of Biological Sciences, Ashworth Laboratories, University of Edinburgh, Edinburgh EH9 3JT, UK.



# BRIEF COMMUNICATIONS ARISING

---

- <sup>24</sup>Department of Psychology, University of California, Santa Barbara, Santa Barbara, California 93106-9660, USA.
- <sup>25</sup>Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08540, USA.
- <sup>26</sup>Department of Ecology and Evolution, The University of Chicago, 1101 E. 57 Street, Chicago, Illinois 60637, USA.
- <sup>27</sup>Department of Ecology, Montana State University, Bozeman, Montana 59717, USA.
- <sup>28</sup>Department of Biosciences, 8888 University Drive, Simon Fraser University, Burnaby, British Columbia V5A1S6, Canada.
- <sup>29</sup>Department of Biology, Villanova University, 800 Lancaster Avenue, Villanova, Pennsylvania 19085, USA.
- <sup>30</sup>Department of Mathematics and Statistics, Queen's University, Kingston, Ontario K7L 3N6, Canada.
- <sup>31</sup>The Cornell Laboratory of Ornithology, Cornell University — The Johnson Center, 159 Sapsucker Woods Road, Ithaca, New York 14850, USA.
- <sup>32</sup>Department of Biology, University of Louisville, Louisville, Kentucky 40292, USA.
- <sup>33</sup>Seeley G. Mudd Hall, Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA.
- <sup>34</sup>USDA-ARS Bee Research Laboratory, BARC-E Bldg 476, Beltsville, Maryland 20705, USA.
- <sup>35</sup>Laboratoire Ecologie and Evolution, CNRS UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France; Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.
- <sup>36</sup>School of Life Sciences, John Maynard Smith Building, University of Sussex, Brighton BN1 9QG, UK.
- <sup>37</sup>Department Biologie II Behavioral Ecology (Verhaltensökologie), Ludwig-Maximilians-Universität, München Großhaderner Str. 2 D - 82152 Planegg/Martinsried, Germany.
- <sup>38</sup>Department of Entomology, University of Kentucky, Lexington, Kentucky 40546-0091, USA.
- <sup>39</sup>School of Life Sciences, Arizona State University, PO Box 874501, Tempe, Arizona 85287-4501, USA.
- <sup>40</sup>CEFE - UMR 5175, 1919 route de Mende, F-34293 Montpellier Cedex 5, France.
- <sup>41</sup>School of Biological Sciences, Flinders University, GPO Box 2100, Adelaide, South Australia 5001, Australia.
- <sup>42</sup>Kellogg Biological Station and Department of Zoology, Michigan State University, Hickory Corners, Michigan 49060, USA.
- <sup>43</sup>School of Biology and Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, 310 Ferst Drive, Atlanta, Georgia 30332-0230, USA.
- <sup>44</sup>Department of Evolution and Ecology, College of Biological Sciences, 1 Shields Avenue, UC Davis, Davis, California 95616, USA.
- <sup>45</sup>Center for Pollinator Research, Huck Institutes of the Life Sciences, Pennsylvania State University, Chemical Ecology Lab 4A, University Park, Pennsylvania 16802, USA.
- <sup>46</sup>Muséum National d'Histoire Naturelle, CP39, 12 rue Buffon, 75005 Paris, France.
- <sup>47</sup>Biology Department, University of Toronto, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada.
- <sup>48</sup>Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.
- <sup>49</sup>Biologie I, Universität Regensburg, D-93040 Regensburg, Germany.
- <sup>50</sup>Department of Biosciences, PL 65 (Viikinkaari 1), FI-00014 University of Helsinki, Finland.
- <sup>51</sup>Department of Biology, University of Vermont, Burlington, Vermont 05405, USA.
- <sup>52</sup>School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona 85287-2402, USA.
- <sup>53</sup>Department of Animal Ecology, Institute of Ecological Science, Faculty of Earth and Life Sciences, Vrije Universiteit, De Boelelaan 1085, NL-1081 HV Amsterdam, The Netherlands.
- <sup>54</sup>Animal Ecology Group, Centre for Evolutionary and Ecological Studies, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands.
- <sup>55</sup>University of Osnabrueck, Barbarastr.11, D-49076 Osnabrueck, Germany.
- <sup>56</sup>Harvard University, Museum of Comparative Zoology, 26 Oxford St, Cambridge, Massachusetts 02138, USA.
- <sup>57</sup>Environmental Microbiology, Swiss Federal Institute of Aquatic Research and Technology, Überlandstrasse 133, CH-8600 Dübendorf, Switzerland.
- <sup>58</sup>Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark.
- <sup>59</sup>School of Biological Sciences, Royal Holloway, University of London, Egham TW20 0EX, UK.
- <sup>60</sup>Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95064, USA.
- <sup>61</sup>Department of Computer Science, University of Sheffield, Sheffield S1 4DP, UK.
- <sup>62</sup>Department of Anthropology and Center for Population Biology, UC Davis, Davis, California 95616, USA.
- <sup>63</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.
- <sup>64</sup>Department of Zoology, 730 Van Vleet Oval, University of Oklahoma, Norman, Oklahoma 73019, USA.
- <sup>65</sup>Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada.
- <sup>66</sup>Integrative Biology, University of Texas at Austin, 1 University Station C0930, Austin, Texas 78712, USA.
- <sup>67</sup>Psychologie — Université de Strasbourg, Ethologie des Primates — DEPE (IPHC CNRS UMR 7178), 23 rue Becquerel — Strasbourg 67087, Cedex, France.
- <sup>68</sup>Department of Philosophy, University of Bristol, Bristol BS8 1TB, UK.
- <sup>69</sup>Biocenter Oulu and Department of Biosciences, University of Helsinki, Box 65, 00140 University of Helsinki, Finland.
- <sup>70</sup>Institute of Integrative Biology, Biosciences Building, Crown Street, University of Liverpool, Liverpool L69 7ZB, UK.
- <sup>71</sup>Theoretical Biology group, University of Groningen, PO Box 14, 9750 AA Haren, The Netherlands.
- <sup>72</sup>Department of Biology, CB#3280, Coker Hall, University of North Carolina, Chapel Hill, NC 27599-3280, USA.
- <sup>73</sup>Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas 77005-1892, USA.
- <sup>74</sup>Department of Biochemistry, University of Zurich, Building Y27, Office J-46, Winterthurststrasse 190, CH-8057 Zurich, Switzerland; Swiss Institute of Bioinformatics, Quartier Sorge Bâtiment Génomode, CH- 1015 Lausanne, Switzerland.
- <sup>75</sup>Research Department of Genetics, Evolution and Environment, Faculty of Life Sciences, University College London, 4 Stephenson Way, London NW1 2HE, UK.
- <sup>76</sup>Centre for Behaviour and Evolution, Institute of Neuroscience, Faculty of Medical Sciences, Newcastle University, Henry Wellcome Building, Framlington Place, Newcastle upon Tyne NE2 4HH, UK.
- <sup>77</sup>School of Marine and Tropical Biology, James Cook University, Queensland 4811, Australia.
- <sup>78</sup>Station Biologique de Roscoff, CNRS-UPMC UMR 7144, 29680 Roscoff, France.
- <sup>79</sup>Institut des Sciences de l'Evolution, University of Montpellier 2, Montpellier 34095, France.
- <sup>80</sup>Department of Biology, University of North Carolina at Greensboro, 312 Eberhart Building, Greensboro, North Carolina 27403, USA.

<sup>81</sup>Department of Biology, 3314 Spieth Hall, University of California — Riverside, Riverside, California 92521, USA.

<sup>82</sup>ETH Zurich, Institute of Integrative Biology (IBZ), Universitätsstrasse 16, CH-8092 Zürich, Switzerland.

<sup>83</sup>School of Philosophy, Psychology and Language Sciences, University of Edinburgh, 3 Charles Street, Edinburgh EH8 9AD, UK.

<sup>84</sup>School of Biology, University of St Andrews, Harold Mitchell Building, St Andrews, Fife KY16 9TH, UK.

<sup>85</sup>William Paterson University of New Jersey, 300 Pompton Road, Wayne, New Jersey 07470, USA.

<sup>86</sup>Department of Anthropology, 101 West Hall, University of Michigan, Ann Arbor, Michigan 48109, USA.

<sup>87</sup>Department of Entomology and Department of Animal Biology, University of Illinois, Urbana, Illinois 61801, USA.

<sup>88</sup>Behavioural Ecology, Institute of Ecology and Evolution, University of Bern, Wohlenstrasse 50a, CH-3032 Hinterkappelen, Switzerland.

<sup>89</sup>Department of Biology, University of Western Ontario, 1151 Richmond Street North, London, Ontario N6A 5B7, Canada.

<sup>90</sup>Department of Anthropology, University of California, Santa Barbara, California 93106-3210, USA.

<sup>91</sup>Department of Environmental Science, Policy and Management, 130 Mulford Hall, 3114, University of California Berkeley, Berkeley, California 94720-3114, USA.

<sup>92</sup>Faculty of Agriculture, University of the Ryukyus, Okinawa 903-0213, Japan.

<sup>93</sup>Dipartimento di Biologia Evoluzionistica, Università degli Studi di Firenze, via Romana 17, 50125 Firenze, Italy.

<sup>94</sup>Department of Ecology and Evolutionary Biology, University of Tennessee Knoxville, Knoxville, Tennessee 37902, USA.

<sup>95</sup>Department of Entomology, Box 7613, North Carolina State University, Raleigh, North Carolina 27695-7613, USA.

<sup>96</sup>Institute for Theoretical Biology, Humboldt University zu Berlin, Invalidenstr. 43, D-10115 Germany.

<sup>97</sup>Department of Biology, Zoological Institute, K.U. Leuven, Naamsestraat 59, B-3000 Leuven, Belgium.

<sup>98</sup>Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Panamá.

<sup>99</sup>Department of Biology, 101 Morgan Building, University of Kentucky, Lexington, Kentucky 40506-0225, USA.

<sup>100</sup>Department of Applied Mathematics, University of Western Ontario, 1151 Richmond Street North, London, Ontario N6A 5B7, Canada.

<sup>101</sup>Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

<sup>102</sup>Department of Biology and Program in Ecology, Evolution and Conservation Biology, University of Nevada, Reno, Nevada 89557, USA.

<sup>103</sup>Department of Biology, San Francisco State University, San Francisco, California 94132, USA.

Received 20 September; accepted 17 December 2010.

- Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
- Hamilton, W. D. The genetical evolution of social behaviour, I & II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Hamilton, W. D. Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220 (1970).
- Grafen, A. Optimisation of inclusive fitness. *J. Theor. Biol.* **238**, 541–563 (2006).
- Frank, S. A. *Foundations of Social Evolution* (Princeton Univ. Press, 1998).
- Gardner, A., West, S. A. & Barton, N. H. The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**, 207–226 (2007).
- Rousset, F. *Genetic Structure and Selection in Subdivided Populations* (Princeton Univ. Press, 2004).
- Queller, D. C. A general model for kin selection. *Evolution* **46**, 376–380 (1992).
- Krebs, J. R. & Davies, N. B. *Behavioural Ecology. An Evolutionary Approach* 4th edn (Blackwell Scientific, 1997).
- Foster, K. R. A defense of sociobiology. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 403–418 (2009).
- Westneat, D. F. & Fox, C. W. *Evolutionary Behavioral Ecology* (Oxford Univ. Press, 2010).
- Hardy, I. C. W. *Sex Ratios: Concepts and Research Methods* (Cambridge Univ. Press, 2002).
- West, S. A. *Sex Allocation* (Princeton Univ. Press, 2009).
- Queller, D. C. & Strassmann, J. E. Kin selection and social insects. *Bioscience* **48**, 165–175 (1998).
- Boomsma, J. J. Lifetime monogamy and the evolution of eusociality. *Phil. Trans. R. Soc. Lond. B* **364**, 3191–3207 (2009).
- Ratnieks, F. L. W., Foster, K. R. & Wenseleers, T. Conflict resolution in insect societies. *Annu. Rev. Entomol.* **51**, 581–608 (2006).

**Author Contributions** All authors contributed to the planning, writing and/or revising of this paper. Several others who contributed significantly are not listed because they are named on separate comments.

**Competing financial interests:** declared none.

doi:10.1038/nature09831

## Only full-sibling families evolved eusociality

ARISING FROM M. A. Nowak, C. E. Tarnita & E. O. Wilson *Nature* **466**, 1057–1062 (2010)

The paper by Nowak *et al.*<sup>1</sup> has the evolution of eusociality as its title, but it is mostly about something else. It argues against inclusive fitness theory and offers an alternative modelling approach that is claimed to be more fundamental and general, but which, we believe, has no practical biological meaning for the evolution of eusociality. Nowak *et al.*<sup>1</sup> overlook the robust empirical observation that eusociality has only arisen in clades where mothers are associated with their full-sibling offspring; that is, in families where the average relatedness of offspring to siblings is as high as to their own offspring, independent of population structure or ploidy. We believe that this omission makes the paper largely irrelevant for understanding the evolution of eusociality.

Eusociality is not just any form of condition-dependent reproductive altruism as found in cooperative breeders, but the permanent division of reproductive labour. Clades where helpers became irreversibly eusocial (ants, some bees, some wasps, and termites<sup>2</sup>) are old, radiated into many subclades over evolutionary time, and achieved considerable ecological footprints. A recent comparative study<sup>3</sup> showed that all hymenopteran

clades that fit the standard definition of eusociality<sup>4</sup> evolved from lifetime monogamous ancestors<sup>5–8</sup>. This implies that high relatedness always preceded or coincided with eusociality, and contrasts with the contention of Nowak *et al.*<sup>1</sup> that eusociality can evolve in any group with parental care, or that high relatedness arises after eusociality.

Given that promiscuity is the most common mating system in animals, strict ancestral monogamy throughout eusocial clades implies that high relatedness was necessary for eusociality to evolve. Nonetheless, necessity does not imply sufficiency. Monogamous lineages may have remained solitary because the benefits of helping at the nest were insufficient to surpass independent breeding. This is elegantly captured by the ratio of the parameters *b* and *c* in Hamilton's rule. In a number of ant, bee and wasp genera the high relatedness condition for eusociality has become secondarily relaxed via evolutionary elaborations such as multiple queen mating, but this has only occurred after worker phenotypes had specialized so that opting out to independent breeding had become selectively disadvantageous or

developmentally impossible<sup>3</sup>. Claiming (in their Supplementary Information, Part B) that it is far simpler to consider that advanced eusocial species just need more sperm<sup>1</sup> muddles proximate and ultimate explanations<sup>9,10</sup>; many multiply-mating queens discard most of the sperm they receive<sup>11,12</sup>, indicating that sperm limitation cannot explain polyandry.

We now also know that departures from high relatedness would almost certainly have prevented the evolution of eusociality if they had happened before sterile castes had become permanent<sup>8</sup>, that is, before reaching the point of no return to breeding independently<sup>13</sup>. A recent comparative study on birds<sup>14</sup> showed that cooperative breeding is an unstable state that predominantly occurs in monogamous clades and is likely to be lost when parents become more promiscuous. This evidence is not merely correlative: differences in ancestral promiscuity between cooperative and non-cooperative species were found even before cooperation arose, illustrating that monogamy preceded the evolution of helping and that helpers leave when relatedness incentives are reduced. This shows that high relatedness among siblings is critical along with the Hamiltonian *b/c* ratio but, as in the insects, relatedness is not sufficient because many monogamous birds are not cooperative breeders.

In light of these reconstructions of the ancestral life histories of numerous social clades, it is surprising that the argument of Nowak *et al.*<sup>1</sup> about eusocial evolution starts by assuming that family structure can be replaced by any form of population structure. This assumption is puzzling given the lack of empirical evidence that this hypothetical ‘parasocial’ route to eusociality<sup>1,4</sup> (where same-generation individuals associate independent of relatedness) has produced a single extant clade with obligately eusocial workers. We believe that this renders Part A of the Supplementary Information of Nowak *et al.*<sup>1</sup>, and the arguments throughout the first two-thirds of the paper, largely irrelevant to the origin of eusociality. Part C of the Supplementary Information addresses the evolution of sterile workers within monogamous or clonal families, meaning that relatedness in these models is invariant. As a consequence, we believe that these models have nothing to say about the importance of relatedness in the evolution of eusociality beyond showing that costs and benefits are also important. This was already clear from Hamilton’s rule nearly half a century ago.

It should give pause for thought that none of the long-recognized approximations of inclusive fitness theory raised in the paper was important enough to preclude kin selection theory from developing into a well-integrated network of complementary hypotheses with high predictive power for reproductive decision-making in real-world social organisms. In contrast, the abstractions of Nowak *et al.*<sup>1</sup> fail to provide any new predictions or questions; all they apparently have to offer is the truism that helpers are associated with longer-lived, fecund breeders.

Jacobus J. Boomsma<sup>1</sup>, Madeleine Beekman<sup>2</sup>, Charlie K. Cornwallis<sup>3</sup>, Ashleigh S. Griffin<sup>3</sup>, Luke Holman<sup>1</sup>, William O. H. Hughes<sup>4</sup>, Laurent Keller<sup>5</sup>, Benjamin P. Oldroyd<sup>2</sup> & Francis L. W. Ratnieks<sup>6</sup>

<sup>1</sup>Centre for Social Evolution, Department of Biology, University of Copenhagen, 2100 Copenhagen, Denmark.

e-mail: JJBoomsma@bio.ku.dk

<sup>2</sup>Behaviour and Genetics of Social Insects Lab, School of Biological Sciences A12, University of Sydney, New South Wales, Australia.

<sup>3</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

<sup>4</sup>Institute of Integrative and Comparative Biology, Miall Building, University of Leeds, Leeds LS2 9JT, UK.

<sup>5</sup>Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland.

<sup>6</sup>Laboratory of Apiculture and Social Insects, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

Received 19 September; accepted 17 December 2010.

- Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
- Inward, D. J. G., Vogler, A. P. & Eggleton, P. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* **44**, 953–967 (2007).
- Hughes, W. O. H., Oldroyd, B. P., Beekman, M. & Ratnieks, F. L. W. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–1216 (2008).
- Wilson, E. O. *The Insect Societies* (Belknap Press of Harvard Univ. Press, 1971).
- Hamilton, W. D. The genetical evolution of social behaviour, I & II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Alexander, R. D. The evolution of social behavior. *Annu. Rev. Ecol. Syst.* **5**, 325–383 (1974).
- Charnov, E. L. Evolution of eusocial behavior: offspring choice or parental parasitism? *J. Theor. Biol.* **75**, 451–465 (1978).
- Boomsma, J. J. Kin selection versus sexual selection: Why the ends do not meet. *Curr. Biol.* **17**, R673–R683 (2007).
- Mayr, E. Cause and effect in biology. *Science* **134**, 1501–1506 (1961).
- Tinbergen, N. On aims and methods of ethology. *Z. Tierpsychol.* **20**, 410–433 (1963).
- Baer, B. Sexual selection in *Apis* bees. *Apidologie (Celle)* **36**, 187–200 (2005).
- den Boer, S. P. A. *et al.* Prudent sperm use by leaf-cutter ant queens. *Proc. R. Soc. Lond. B* **276**, 3945–3953 (2009).
- Wilson, E. O. One giant leap: How insects achieved altruism and colonial life. *Bioscience* **58**, 17–25 (2008).
- Cornwallis, C. K., West, S. A., Davis, K. E. & Griffin, A. S. Promiscuity and the evolutionary transition to complex societies. *Nature* **466**, 969–972 (2010).

**Author Contributions** J.J.B. took the initiative for this contribution and wrote the first draft. All co-authors provided written and/or oral comments that helped shape the final submission.

**Competing financial interests:** declared none.

doi:10.1038/nature09832

## Kin selection and eusociality

ARISING FROM M. A. Nowak, C. E. Tarnita & E. O. Wilson *Nature* **466**, 1057–1062 (2010)

Hamilton<sup>1</sup> described a selective process in which individuals affect kin (kin selection), developed a novel modelling strategy for it (inclusive fitness), and derived a rule to describe it (Hamilton’s rule). Nowak *et al.*<sup>2</sup> assert that inclusive fitness is not the best modelling strategy, and also that its production has been ‘meagre’. The former may be debated by theoreticians, but the latter is simply incorrect. There is abundant evidence to demonstrate that inclusive fitness, kin selection and Hamilton’s rule have been extraordinarily productive for understanding the evolution of sociality.

Below we list a few examples of what has been learned from applying kin selection theory—there are thousands of others. (1) Organisms overwhelmingly direct costly assistance, and all true altruism, towards kin<sup>3</sup>. (2) Eusociality in insects originated in organisms with parental care and single mating, which means that relatedness among helpers and brood is generally at the level of siblings<sup>4</sup>. (3) Benefits that can make helping more profitable than reproducing independently often take the forms of either fortress defence (termites, naked mole rats, social shrimp, social thrips and aphids, and some ants) or life insurance



(wasps, bees, other ants)<sup>5,6</sup>. (4) Sex ratios, worker egg laying, worker policing, caste conflict and other social interactions in eusocial insects are explained by kin selection theory<sup>5,7</sup>.

Kin selection evidence is not merely correlative. Numerous kin selection experiments manipulate relatedness in animals, plants and microbes. Other experiments manipulate costs and benefits and show that kin selection is predictive<sup>8</sup>.

Nowak *et al.*<sup>2</sup> say that the haplodiploid hypothesis is not the only explanation for eusociality, but that has not been in dispute for some time. Haplodiploidy is not necessary for the evolution of eusociality, and is not the same as kin selection<sup>5,7</sup>.

Kin selection does not explain all social behaviour, but the claim that it does has never been widely accepted. There are cooperative acts that benefit the actor directly, and between-species mutualisms that must have direct benefits to evolve<sup>9</sup>. But only kin selection can explain true altruism<sup>10</sup>.

Kin selection theory is still inspiring new research. Application of kin selection theory to microbes, including those causing human diseases, is expanding<sup>11</sup>. Kin selection is changing our views of imprinting<sup>12</sup> and maternal–fetus diseases in humans<sup>12</sup>.

Clearly kin selection is a strong, vibrant theory that is the basis for understanding how social behaviour has evolved. Perhaps the best examples come from kin recognition, a field that did not exist before Hamilton's insights<sup>7</sup>. We are puzzled why Nowak *et al.*<sup>2</sup> would attack a body of research that has been exemplary as “a domain of empirical knowledge [that has] followed so closely and fruitfully upon an abstract theoretical idea.”<sup>13</sup>.

**Joan E. Strassmann<sup>1</sup>, Robert E. Page Jr<sup>2</sup>, Gene E. Robinson<sup>3</sup> & Thomas D. Seeley<sup>4</sup>**

<sup>1</sup>Department of Ecology and Evolutionary Biology, Rice University, Houston, Texas, USA.  
e-mail: strassm@rice.edu

<sup>2</sup>School of Life Sciences, Arizona State University, Tempe, Arizona, USA.

<sup>3</sup>Department of Entomology, and Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.

<sup>4</sup>Department of Neurobiology and Behavior, Cornell University, Ithaca, New York, USA.

**Received 19 November; accepted 17 December 2010.**

1. Hamilton, W. D. The genetical evolution of social behaviour. I–II. *J. Theor. Biol.* **7**, 1–52 (1964).
2. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
3. Griffin, A. S. & West, S. A. Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science* **302**, 634–636 (2003).
4. Hughes, W., Oldroyd, B., Beekman, M. & Ratnieks, F. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–1216 (2008).
5. Queller, D. C. & Strassmann, J. E. Kin selection and social insects. *Bioscience* **48**, 165–175 (1998).
6. Strassmann, J. & Queller, D. Insect societies as divided organisms: the complexities of purpose and cross-purpose. *Proc. Natl Acad. Sci. USA* **104**, 8619–8626 (2007).
7. Crozier, R. H. & Pamilo, P. *Evolution of Social Insect Colonies: Sex Allocation and Kin Selection* (Oxford Univ. Press, 1996).
8. Field, J. & Cant, M. A. Social stability and helping in small animal societies. *Phil. Trans. R. Soc. B* **364**, 3181–3189 (2009).
9. Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160 (2004).
10. Lehmann, L. & Keller, L. The evolution of cooperation and altruism. A general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376 (2006).
11. West, S. A., Diggle, S. P., Buckling, A., Gardner, A. & Griffins, A. S. The social lives of microbes. *Annu. Rev. Ecol. Evol. Syst.* **38**, 53–77 (2007).
12. Haig, D. Evolution of placental hormones. *Am. J. Hum. Biol.* **20**, 222–223 (2008).
13. Wilson, E. O. In *Kin Recognition in Animals* (eds Fletcher, D. J. C. & Michener, C. D.) 7–18 (Wiley, 1987).

**Author Contributions** All authors conceived the project, analysed the material and wrote the paper.

**Competing financial interests:** declared none.

doi:10.1038/nature09833

## Inclusive fitness in evolution

ARISING FROM M. A. Nowak, C. E. Tarnita & E. O. Wilson *Nature* **466**, 1057–1062 (2010)

For over fifty years, the evolution of social behaviour has been guided by the concept of inclusive fitness as a measure of evolutionary success. Nowak *et al.*<sup>1</sup> argue that inclusive fitness should be abandoned. In so doing, however, they misrepresent the role that inclusive fitness has played in the theory of social evolution by which understanding social behaviour in a variety of disciplines has developed and flourished. By discarding inclusive fitness on the basis of its limitations, they create a conceptual tension which, we argue, is unnecessary, and potentially dangerous for evolutionary biology.

The core argument of Nowak *et al.*<sup>1</sup> for abandoning inclusive fitness is its limited capacity to predict dynamics in evolutionary models. This is an old point, and one that was hotly debated in the early years of kin selection theory<sup>2</sup>. Inclusive fitness was developed by Hamilton to summarize a difficult frequency-dependent selection problem by using a simple maximization principle. Early work<sup>3</sup> proved that the average inclusive fitness effect is maximized by behavioural evolution in family structured populations and that it provides the surface for the Wright's adaptive topography (arguably one of the most useful tools that has ever been developed for understanding evolution).

Hamilton's great insight was that individual fitness is not maximized by social evolution when relatives are present, inclusive fitness is. The

idea that something other than the individual organism could be the fitness-maximizing unit was completely revolutionary at the time and opened new research areas that are still being developed, such as the study of transitions in units of evolution and individuality<sup>4</sup>.

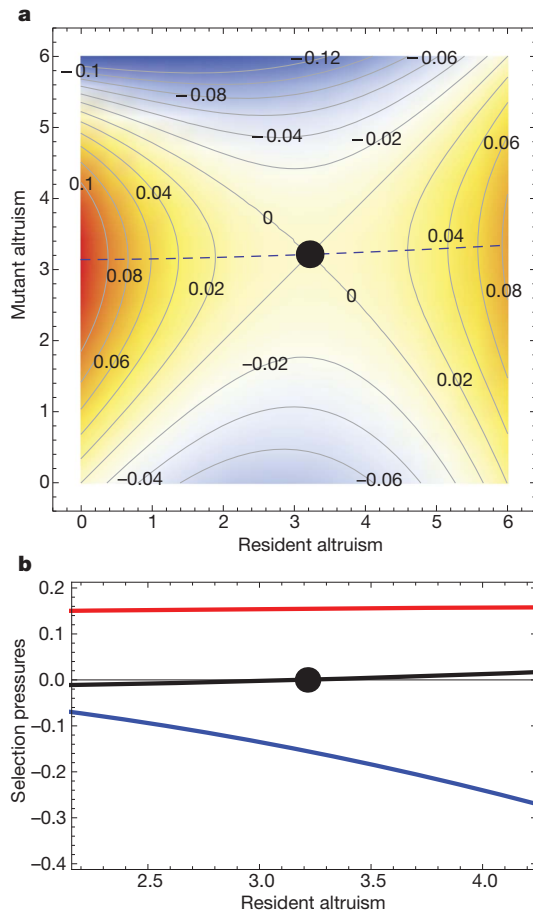
Today, inclusive fitness and evolutionary dynamics models are bridged and linked by the unifying concept of invasion fitness<sup>5–11</sup> (Fig. 1). Invasion fitness captures feedbacks between the evolution of social traits and the ecological structure of the evolving population<sup>7–11</sup> (Fig. 2). Invasion fitness embraces average fitness and inclusive fitness maximization at evolutionary steady states<sup>5,7,9–11</sup>, and further reveals the difficulty of reducing evolution to a simple maximization process<sup>5,12</sup> (Fig. 1a).

Under Hamilton's rule, the condition for invasion of an altruistic allele involves a linear function of genetic relatedness. With eco-evolutionary feedbacks, Hamilton's rule becomes part of a more complex and dynamic framework<sup>7,10,11</sup> in which relatedness between interacting individuals is a dynamic property of the population and an outcome of the models, rather than a pre-defined feature (which relatedness was in earlier models of inclusive fitness). Yet in this new framework the assumption of weak selection alone is often sufficient for Hamilton's rule to predict accurately endpoints of altruism

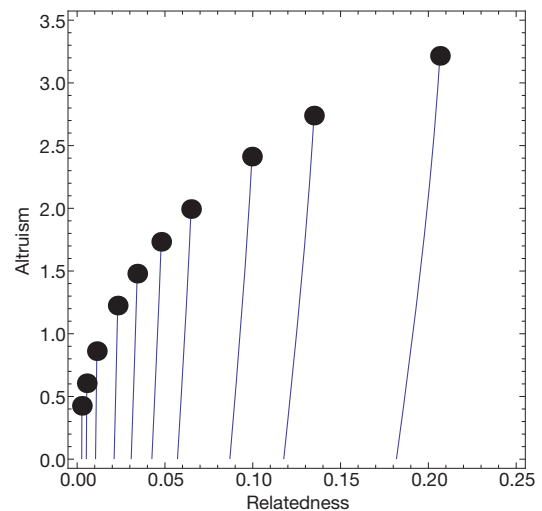
# BRIEF COMMUNICATIONS ARISING

evolution<sup>11</sup> (Fig. 1b). This framework has also begun to provide the testable predictions under competing hypotheses about relatedness and sociality that Nowak *et al.*<sup>1</sup> call for<sup>8–11</sup> (Fig. 2).

Both inclusive fitness and average fitness maximization are general insights about the evolutionary process with great heuristic value, even though they rely on special conditions to predict evolutionary dynamics. Considerable progress in extending evolutionary dynamics models to more general ecological, behavioural and genetic scenarios has been guided by the inclusive fitness concept<sup>4,6–11,13–15</sup>. By opposing ‘standard selection theory’ and ‘inclusive fitness theory’, we believe that Nowak *et al.*<sup>1</sup> give the incorrect (and potentially dangerous) impression that evolutionary thinking has branched out into conflicting and apparently incompatible directions. In fact, there is only one paradigm: natural selection driven by interactions, interactions of all kinds and at all levels. Inclusive fitness has been a powerful force in the development of this paradigm and is likely to have a continued role in the evolutionary theory of behaviour interactions.



**Figure 1 | Fitness landscape and Hamilton's rule.** **a**, Fitness landscape in resident–mutant phenotypic space. Mutant invasion (inclusive) fitness is maximized along the dashed curve and the black circle is an evolutionarily stable phenotype. However, the asymmetrical sign structure implies that evolutionary dynamics do not obey an optimizing principle<sup>12</sup>. **b**, Selection pressures arise from relatedness (red curve), physiological cost of altruism (blue) and change in space occupancy (black). Under weak selection, the latter is negligible (black curve close to zero) and the benefit is proportional to relatedness (see for example, equation (3) in ref. 10), which makes Hamilton's rule a good approximation of selection<sup>10,11</sup>. The black circle indicates the evolutionarily stable phenotype found in **a**.



**Figure 2 | Co-evolution of altruism and relatedness.** Evolutionary trajectories (thin curves) start with zero altruism and converge to evolutionarily stable phenotypes (black circles). Relatedness is dynamic and co-evolves with altruism. Eco-evolutionary models can predict how environmental factors (for example, habitat viscosity) affect altruism evolution. Here, viscosity decreases across trajectories from right to left, and the string of black circles shows how evolved altruism and relatedness should co-vary along an environmental gradient of habitat viscosity.

## METHODS

See ref. 10 for details on the model underlying Figs 1 and 2 (and others<sup>7,8,11</sup> for further development). Altruism is a continuous character (with haploid inheritance) evolving in a spatial population network. Invasion fitness is the growth rate of a self-structured mutant cluster. Figure 1a, b is based, respectively, on figures 6E and 4E in ref. 10 with mobility rate = 1 and habitat viscosity = 1/4. Figure 1a makes no special assumption on selection, interactions or population structure. Figure 2 is based on figures 5F and 7F in ref. 10 with mobility rate = 1 and habitat viscosity from 1/4 (right) to 1/256 (left). Figures 1b and 2 only assume weak selection.

Regis Ferriere<sup>1,2</sup> & Richard E. Michod<sup>2</sup>

<sup>1</sup>Laboratoire Ecologie & Evolution, CNRS UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France.

e-mail: regis.ferriere@ens.fr

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA.

Received 27 September; accepted 17 December 2010.

- Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
- Michod, R. E. The theory of kin selection. *Annu. Rev. Ecol. Syst.* **13**, 23–55 (1982).
- Michod, R. E. & Abugov, R. Adaptive topography in family structured models of kin selection. *Science* **210**, 667–669 (1980).
- Michod, R. E. *Darwinian Dynamics* (Princeton Univ. Press, 1999).
- Metz, J. A. J. in *Encyclopedia of Ecology* (eds Jorgensen, S. E. & Fath, B. D.) (Elsevier, 2008).
- Roff, D. A. *Modeling Evolution* (Oxford Univ. Press, 2010).
- van Baalen, M. & Rand, D. A. The unit of selection in viscous populations and the evolution of altruism. *J. Theor. Biol.* **193**, 631–648 (1998).
- Ferriere, R. & Le Galliard, J.-F. in *Dispersal* (eds Clobert, J., Danchin, E., Dhondt, A. A. & Nichols, J. D.) 57–79 (Cambridge Univ. Press, 2001).
- Cadet, C., Ferriere, R., Metz, J. A. J. & van Baalen, M. The evolution of dispersal under demographic stochasticity. *Am. Nat.* **162**, 427–441 (2003).
- Le Galliard, J.-F., Ferriere, R. & Dieckmann, U. The adaptive dynamics of altruism in spatially heterogeneous populations. *Evolution* **57**, 1–17 (2003).
- Lion, S. & van Baalen, M. From infanticide to parental care: why spatial structure can help adults be good parents. *Am. Nat.* **170**, E26–E46 (2007).
- Metz, J. A. J., Mylius, S. D. & Dieckmann, O. When does evolution optimize? *Evol. Ecol. Res.* **10**, 629–654 (2008).
- Michod, R. E. & Sanderson, M. in *Evolution—Essays in Honour of John Maynard Smith* Ch. 7 (eds Greenwood, P. J. & Slatkin, M.) 95–104 (Cambridge Univ. Press, 1985).

14. Day, T. & Taylor, P. D. Unifying genetics and game theoretic models of kin selection for continuous traits. *J. Theor. Biol.* **194**, 391–407 (1998).
15. Rousset, F. Separation of time scales, fixation probabilities and convergence to evolutionarily stable states under isolation by distance. *Theor. Popul. Biol.* **69**, 165–179 (2006).

**Author Contributions** R.F. and R.E.M. conceived the study. R.F. analysed the model and prepared the figures. R.F. and R.E.M. wrote the paper.

**Competing financial interests:** declared none.

doi:10.1038/nature09834

## In defence of inclusive fitness theory

ARISING FROM M. A. Nowak, C. E. Tarnita & E. O. Wilson *Nature* **466**, 1057–1062 (2010)

Arguably the defining characteristic of the scientific process is its capacity for self-criticism and correction<sup>1</sup>. Nowak *et al.*<sup>2</sup> challenge proposed connections between relatedness and the evolution of eusociality<sup>3</sup>, suggest instead that defensible nests and “spring-loaded” traits are key, and present alternative modelling approaches. They then dismiss the utility of Hamilton’s insight that relatedness has a profound evolutionary effect<sup>3</sup>, formalized in his widely accepted inclusive fitness theory as Hamilton’s rule (“Rise and fall of inclusive fitness theory”). However, we believe that Nowak *et al.*<sup>2</sup> fail to make their case for logical, theoretical and empirical reasons.

Logically, both in attacking inclusive fitness and in attempting to reinforce their own positions, Nowak *et al.*<sup>2</sup> cherry-pick examples and fail to distinguish necessary from sufficient causes<sup>1</sup>. Yes, there are hundreds of haplodiploid species that are not eusocial<sup>2</sup>. Yet, there are also hundreds of nest-making (diploid) birds, mammals and reptiles that are not eusocial. Moreover, if the non-eusocial, haplodiploid species pose a problem for inclusive fitness, then the fact that hundreds of them also make nests (including many living in communal or sub-social groups) does not support the proposed alternative.

Theoretically, in promoting their modelling approach, Nowak *et al.*<sup>2</sup> pose a false dichotomy between inclusive fitness theory and “standard natural selection theory”. They assert, we believe incorrectly, that inclusive fitness theory suffers from numerous ills (for example, “stringent assumptions”), yet their own models require stringent assumptions, without the benefit of any generality. Indeed, although asserting that “relatedness does not drive the evolution of eusociality”, the authors do not present the critical test of removing the effects of relatedness in their model (for example, by randomly assigning daughters to nests). Thus, Nowak *et al.*<sup>2</sup> do not provide any basis for their core assertion, and available data on real biological systems<sup>4–6</sup> directly contradict it.

Empirically, Nowak *et al.*<sup>2</sup>, in our eyes, misinterpret relevant literature. Emphasizing progressive provisioning of food to immatures as a critical pre-adaptation (that is, a “spring-loaded” trait), they overlook taxa (for example, sweat bees) in which eusociality evolved repeatedly without progressive provisioning<sup>7,8</sup>. It has been suggested that eusociality might rapidly evolve<sup>9</sup>, but the statement by Nowak *et al.*<sup>2</sup> that studies of forced sociality in *Lasioglossum* bees show that solitary bees will divide labour “in foraging, tunnelling, and guarding” is incorrect. *Lasioglossum hemichalceum* is social (communal), not solitary<sup>10</sup>, and the solitary *Lasioglossum figueresi* was studied in artificial arenas, not nests, so it was impossible for bees to forage, tunnel or guard<sup>11</sup>. Moreover, the small carpenter bees that Nowak *et al.*<sup>2</sup> cite are in a genus (*Ceratina*) that contains no known obligately eusocial species, and only one species in which facultative eusociality occurs at high frequency<sup>12</sup>, indicating that even if “spring-loaded” traits exist, Nowak *et al.*<sup>2</sup> have misidentified them.

What is clear is that neither haplodiploidy, nests, nor “spring-loaded” traits is sufficient for the evolution of eusociality. However, the most recent comparative evidence supports the basic prediction of

inclusive fitness theory that, regardless of ploidy or the presence of nests or “spring-loaded” traits, high relatedness is key to the evolution of cooperative breeding and/or eusociality<sup>4–6</sup>. Any serious attempt to dismiss inclusive fitness theory must address the results of these important comparative studies<sup>4–6</sup> directly.

Beyond its being completely integrated with “standard natural selection theory”<sup>13</sup>, beyond extensive theoretical work showing that it is both flexible and robust<sup>13</sup>, beyond the fact that available evidence supports its fundamental prediction that high relatedness is key for the evolution of eusociality<sup>4–6</sup>, inclusive fitness theory has the virtue of making general, non-obvious predictions well beyond the issue of eusociality<sup>4–6</sup>. Kin recognition and policing<sup>14,15</sup>, mother–fetus conflicts, and patterns of sex allocation (particularly in eusocial insects) stand out<sup>3,14,15</sup>. Collectively, those predictions have again and again been borne out in a vast comparative and experimental empirical literature (for example, refs 3–6, 14, 15) that Nowak *et al.*<sup>2</sup> nonetheless dismiss as “meagre” and “superficial”. Nowak *et al.*<sup>2</sup> present a provocative essay, but in their apparent rush to discard inclusive fitness theory, they present an alternative that we believe to be deeply flawed. Although the continued scrutiny of accepted paradigms is an essential part of the scientific process, the reports<sup>2</sup> of the fall of inclusive fitness theory have been greatly exaggerated. If anything, Nowak *et al.*<sup>2</sup> succeed in reminding us of the elegance and power of Hamilton’s numerous insights and contributions<sup>3</sup>.

Edward Allen Herre<sup>1</sup> & William T. Wcislo<sup>1</sup>

<sup>1</sup>Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Republic of Panama; Smithsonian Tropical Research Institute, MRC 0580-06, unit 9100 Box 0948, DPO AA 34002-9998, USA. e-mail: herrea@si.edu

Received 8 October; accept 17 December 2010.

1. Mayr, E. *The Growth of Biological Thought* (Harvard Univ. Press, 1982).
2. Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
3. Hamilton, W. D. *Narrow Roads of Gene Land*. Vol. I. Evolution of Social Behaviour (Oxford Univ. Press, 1998).
4. Hughes, W. O. H., Oldroyd, B. P., Beekman, M. & Ratnieks, F. L. W. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–1216 (2008).
5. Cornwallis, C. K., West, S. A., Davis, K. E. & Griffin, A. S. Promiscuity and the evolution to complex societies. *Nature* **466**, 969–972 (2010).
6. Boomsma, J. J. Lifetime monogamy and the evolution of eusociality. *Phil. Trans. R. Soc. B* **364**, 3191–3207 (2009).
7. Costa, J. T. *The Other Insect Societies* (Harvard Univ. Press, 2006).
8. Michener, C. D. *The Social Behavior of the Bees* (Harvard Univ. Press, 1974).
9. Michener, C. D. From solitary to eusocial: Need there be a series of intervening species? *Fortschr. Zool.* **31**, 293–306 (1985).
10. Jeanson, R., Kukul, P. F. & Fewell, J. H. Emergence of division of labour in halictine bees: Contributions of social interactions and behavioural variance. *Anim. Behav.* **70**, 1183–1193 (2005).
11. Wcislo, W. T. Social interactions and behavioral context in a largely solitary bee, *Lasioglossum (Dialictus) figueresi* (Hymenoptera, Halictidae). *Insectes Soc.* **44**, 199–208 (1997).



12. Cronin, A. L. A molecular phylogeny and social behaviour of Japanese *Ceratina* (Hymenoptera, Apidae, Xylocopinae). *Insect Syst. Evol.* **35**, 137–146 (2004).
13. Gardner, A., West, S. A. & Barton, N. H. The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**, 207–226 (2007).
14. Ratnieks, F. L. W. & Vissher, P. K. Worker policing in honeybees. *Nature* **342**, 796–797 (1989).
15. Mueller, U. G. Haplodiploidy and the evolution of facultative sex ratios in a primitively eusocial bee. *Science* **254**, 442–444 (1991).

**Author Contributions** Both authors contributed extensively to all aspects of this work.

**Competing financial interests:** declared none.

doi:10.1038/nature09835

## Nowak *et al.* reply

**REPLYING TO:** P. Abbot *et al.* *Nature* **471**, doi:10.1038/nature09831 (2011); J. J. Boomsma *et al.* *Nature* **471**, doi:10.1038/nature09832 (2011); J. E. Strassmann *et al.* *Nature* **471**, doi:10.1038/nature09833 (2011); R. Ferriere & R. E. Michod *Nature* **471**, doi:10.1038/nature09834 (2011); E. A. Herre & W. T. Wcislo *Nature* **471**, doi:10.1038/nature09835 (2011)

Our paper challenges the dominant role of inclusive fitness theory in the study of social evolution<sup>1</sup>. We show that inclusive fitness theory is not a constructive theory that allows a useful mathematical analysis of evolutionary processes. For studying the evolution of cooperation or eusociality we must instead rely on evolutionary game theory or population genetics. The authors of the five comments<sup>2–6</sup> offer the usual defence of inclusive fitness theory, but do not take into account our new results.

The definition of inclusive fitness given by Hamilton<sup>7</sup> is as follows:

“Inclusive fitness may be imagined as the personal fitness which an individual actually expresses in its production of adult offspring as it becomes after it has been first stripped and then augmented in a certain way. It is stripped of all components which can be considered as due to the individual’s social environment, leaving the fitness which he would express if not exposed to any of the harms or benefits of that environment. This quantity is then augmented by certain fractions of the quantities of harm and benefit which the individual himself causes to the fitnesses of his neighbours. The fractions in question are simply the coefficients of relationship appropriate to the neighbours whom he affects: unity for clonal individuals, one-half for sibs, one-quarter for half-sibs, one-eighth for cousins,...and finally zero for all neighbours whose relationship can be considered negligibly small.”

The concept of inclusive fitness assumes that the fitness of individuals can be split into additive components caused by individual actions. This approach rests on specific assumptions, which need not hold for any particular evolutionary process. Therefore inclusive fitness theory is not a general description of natural selection. In Part A of our Supplementary Information<sup>1</sup> we provide a mathematical analysis to prove this point. If there are non-zero selection intensities, or if there are synergistic interactions, or if there is complex population structure, then it is easy to find situations where personal fitness cannot be partitioned into additive components as needed by inclusive fitness theory. Essentially, inclusive fitness theory requires fitness to be a linear function of individual actions, but a full understanding of social evolution must take into account the nonlinearity inherent in biological systems.

We distinguish between inclusive fitness theory and standard natural selection theory, because the latter does not require fitness to be split into additive components. We have shown that inclusive fitness theory is a proper subset of the standard theory and makes no independent predictions. Any effect of relatedness is fully captured by the standard approach.

Hamilton’s rule states that cooperation can evolve if relatedness exceeds the cost to benefit ratio. If cost and benefit are parameters of individual actions then this rule almost never holds<sup>1,8,9</sup>. There are attempts to make Hamilton’s rule work by choosing generalized cost and benefit parameters<sup>10</sup>, but these parameters are no longer properties of individual phenotypes. They depend on the entire system including

population structure. These extended versions of Hamilton’s rule have no explanatory power for theory or experiment<sup>11</sup>.

Neither inclusive fitness theory nor any formulation of Hamilton’s rule can deal with evolutionary dynamics<sup>12</sup>. This fact alone invalidates the claim that inclusive fitness theory “is as general as the genetical theory of natural selection”<sup>2</sup>.

Several aspects of our paper are misrepresented in the comments<sup>2–6</sup>. One, we do not argue that relatedness is unimportant. Relatedness is an aspect of population structure, which affects evolution<sup>13</sup>. Two, we do not dispute the importance of kin recognition. Conditional behaviour based on kin recognition can be seen as a mechanism for the evolution of cooperation<sup>14</sup>. Three, Part A of our Supplementary Information<sup>1</sup> is not a model for evolution of eusociality, but a mathematical framework that demonstrates the limitations of inclusive fitness theory. Four, Part C of our Supplementary Information<sup>1</sup> provides a mathematical model for the evolution of eusociality, which makes simple and testable predictions and explains the rarity of the phenomenon. Five, monogamy and sex ratio manipulation may be important for the evolution of eusociality; such ideas are best tested in the context of the explicit model that we propose.

Abbot *et al.*<sup>2</sup> claim that inclusive fitness theory has been tested in a large number of biological contexts, but in our opinion this is not the case. We do not know of a single study where an exact inclusive fitness calculation was performed for an animal population and where the results of this calculation were empirically evaluated. Fitting data to generalized versions of Hamilton’s rule is not a test of inclusive fitness theory, which is not even needed to derive such rules.

The limitations of inclusive fitness theory are also demonstrated by its inability to provide useful calculations for microbial evolution<sup>15,16</sup>.

Herre and Wcislo<sup>6</sup> have presented a one-sided account of cases in halictid eusociality, the details of which do not detract in the least from our argument. Halictid bees were not ignored as stated; we cited them three times. Furthermore, communal halictid bees are ‘social’ only in a primitive sense. They occupy a commons-like tunnel but build and defend their own personal cells as solitary bees<sup>17</sup>. Herre and Wcislo<sup>6</sup> point out that the experiments of Wcislo<sup>18</sup> were designed not to allow foraging, tunnelling, or guarding, but do not mention that these behaviours were tested in other experiments<sup>19,20</sup>. Bees are mass provisioners, as Herre and Wcislo<sup>6</sup> say, and we should have used the phrase ‘defence and care of young with mass provisioning (bees) or progressive provisioning (others)’. We thank Herre and Wcislo<sup>6</sup> for pointing out this oversight. Primitively eusocial halictids nevertheless devote considerable care to the cells, guarding them and in many cases opening them to clean out waste.

Various authors mention sex ratio theory, which we do not study in our paper. Nevertheless a precise understanding of sex ratio evolution is based on population genetics and does not require inclusive fitness theory.

There is no support for the claim that evolution maximizes inclusive fitness. Nobody has offered a mathematical statement explaining what should be maximized and for which process.

Hamilton's work has stimulated much empirical research and has led to many measurements of relatedness. But we have shown that we cannot rely on inclusive fitness theory to describe how interactions among related individuals affect evolution. Inclusive fitness theory is neither useful nor necessary to explain the evolution of eusociality or other phenomena. It is time for the field of social evolution to move beyond the limitations of inclusive fitness theory.

**Martin A. Nowak<sup>1</sup>, Corina E. Tarnita<sup>1</sup> & Edward O. Wilson<sup>2</sup>**

<sup>1</sup>Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

e-mail: martin\_nowak@harvard.edu

<sup>2</sup>Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

- Nowak, M. A., Tarnita, C. E. & Wilson, E. O. The evolution of eusociality. *Nature* **466**, 1057–1062 (2010).
- Abbot, P. *et al.* Inclusive fitness theory and eusociality. *Nature* **471**, doi:10.1038/nature09831 (2011).
- Boomsma, J. J. *et al.* Only full-sibling families evolved eusociality. *Nature* **471**, doi:10.1038/nature09832 (2011).
- Strassmann, J. E., Page, R. E. Jr, Robinson, G. E. & Seeley, T. D. Kin selection and eusociality. *Nature* **471**, doi:10.1038/nature09833 (2011).
- Ferriere, R. & Michod, R. E. Inclusive fitness in evolution. *Nature* **471**, doi:10.1038/nature09834 (2011).
- Herre, E. A. & Wcislo, W. T. In defence of inclusive fitness theory. *Nature* **471**, doi:10.1038/nature09835 (2011).
- Hamilton, W. D. The genetical evolution of social behaviour, I, II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Cavalli-Sforza, L. L. & Feldman, M. W. Darwinian selection and “altruism”. *Theor. Popul. Biol.* **14**, 268–280 (1978).
- Karlin, S. & Matessi, C. Kin selection and altruism. *Proc. R. Soc. Lond. B* **219**, 327–353 (1983).
- Queller, D. C. A general model for kin selection. *Evolution* **46**, 376–380 (1992).
- Chuang, J. S., Rivoire, O. & Leibler, S. Cooperation and Hamilton's rule in a simple synthetic microbial system. *Mol. Syst. Biol.* **6**, 398 (2010).
- Traulsen, A. Mathematics of kin- and group-selection: Formally equivalent? *Evolution* **64**, 316–323 (2010).
- Nowak, M. A., Tarnita, C. E. & Antal, T. Evolutionary dynamics in structured populations. *Phil. Trans. R. Soc. B* **365**, 19–30 (2010).
- Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
- Diggle, S. P., Griffin, A. S., Campell, G. S. & West, S. A. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**, 411–414 (2007).
- Smith, J., van Dyken, J. D. & Zee, P. C. A generalization of Hamilton's rule for the evolution of microbial cooperation. *Science* **328**, 1700–1703 (2010).
- Michener, C. D. *The Social Behavior of the Bees* (Harvard Univ. Press, 1974).
- Wcislo, W. T. Social interactions and behavioral context in a largely solitary bee, *Lasioglossum (Dialictus) figueresi* (Hymenoptera, Halictidae). *Insectes Soc.* **44**, 199–208 (1997).
- Jeanson, R., Kukul, P. F. & Fewell, J. H. Emergence of division of labour in halictine bees: Contributions of social interactions and behavioural variance. *Anim. Behav.* **70**, 1183–1193 (2005).
- Sakagami, S. F. & Maeta, Y. Sociality, induced and/or natural, in the basically solitary small carpenter bees (*Ceratina*). In *Animal Societies: Theories and Facts* (eds Itô, Y., Brown, J. L. & Kikkawa, J.) 1–16 (Japan Scientific Societies Press, 1987).

**Author Contributions** M.A.N., C.E.T. and E.O.W. collaborated on all aspects of this reply.

**Competing financial interests:** declared none.

doi:10.1038/nature09836

# natureOUTLOOK

## CANCER PREVENTION

24 March 2011 / Vol 471 / Issue No. 7339



COVER ART:  
MATTHEW HOLLISTER

### Editorial

Herb Brody,  
Michelle Grayson,  
Tony Scully

### Art & Design

Wes Fernandes, Fern  
Bale, Paul Jackman,  
Barbara Izdebska,  
Anthea Lewis

### Production

Donald McDonald,  
Emilia Orviss,  
Daniel Foley

### Sponsorship

Reya Silao, Yvette  
Smith, David Bagshaw

### Marketing

Elena Woodstock,  
Hannah Phipps

### Project Managers

Claudia Deasy, Helen  
Anthony

### Magazine Editor

Tim Appenzeller

### Editor-in-Chief

Philip Campbell

### Consultant editors

Lisa Hutchinson,  
Barbara Marte, Nicola  
McCarthy

Ask many people what they are afraid of, and cancer — the big C — will often top the list. Some forms of cancer have become easily treatable. But in many cases, by the time doctors deploy the weapons of surgery, chemotherapy or radiation, the cancer has already progressed past the point where medical intervention can cure the condition.

Many cancer specialists now contend that the best way to deal with cancer is to ensure it doesn't develop in the first place. Just as the most effective way to deal with frostbite is not to amputate a toe but to wear warm socks and shoes beforehand, the same is true for cancer — the best solution is to enhance our defences and, as best we can, avoid carcinogens.

A small subset of the cancer research community is focusing its attention on this goal. Chemopreventive drugs and vaccines could fortify our healthy bodies against future malignancies (see articles on page S5 and S8). New technologies are revealing precancerous lesions when they can be easily snuffed out (S14). Certain foods with anticancer properties can form part of a healthy diet (S22), while their prophylactic ingredients can be extracted or synthesized.

While curing a patient is a heart-warming triumph; prevention merely maintains the status quo of good health. No surprise which activity gets the lion's share of media attention — and funding. This *Nature Outlook* showcases the progress being made in cancer prevention despite its status as the poor relation of the cancer research world.

We can only hope that the efforts reported here will make cancer as obsolete as the pock marks that not so long ago marred the faces of those who survived smallpox. We are pleased to acknowledge the financial support of the Janssen Pharmaceutical Companies in producing this *Outlook*. As always, *Nature* retains sole responsibility for all editorial content.

**Herb Brody**

*Supplements Editor, Nature Outlook.*

## CONTENTS

### S2 INTRODUCTION

#### **The prevention agenda**

Addressing the research deficit

### S5 CHEMOPREVENTION

#### **First line of defence**

Using drugs to stop cancer in its tracks

### S8 VACCINES

#### **Know your enemy**

Unmasking the cancerous culprits

### S10 PERSPECTIVE

#### **The big C — for Chemoprevention**

Michael B. Sporn

### S12 EPIGENETICS

#### **Unravelling the cancer code**

Exploiting cancer's genetic complexity

### S14 EARLY DETECTION

#### **Spotting the first signs**

Latest tools for the earliest diagnosis

### S16 LIFESTYLE

#### **Breaking the cancer habit**

Healthy living to avoid the disease

### S18 PERSPECTIVE

#### **Tackling the real issues**

Stephen S. Hecht

### S19 BIOMARKERS

#### **Portents of malignancy**

Piecing together the puzzling clues

### S22 FOOD

#### **The omnivore's labyrinth**

Eating your way out of cancer danger

## COLLECTION

**S26 Chemoprevention of colorectal cancer by targeting APC-deficient cells for apoptosis**  
*L. Zhang et al.*

**S30 Molecular targets of phytochemicals for cancer prevention**  
*K.W. Lee, A. M. Bode & Z. Dong*

**S37 Statins lose prevention ground**

**S38 Will an aspirin a day keep the colorectal cancer away?**  
*M. Berg & K. Søreide*

**S40 Evolving role of 5- $\alpha$  reductase inhibitors in chemoprevention**  
*W. M. White & E. D. Kim*

**S42 Fruit and vegetables and cancer risk**  
*T. J. Key*

*Nature Outlooks* are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* Editorial guidelines available at [http://www.nature.com/advertising/resources/pdf/outlook\\_guidelines.pdf](http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf)

### CITING THE OUTLOOK

Cite as a supplement to *Nature*, for example, *Nature* Vol XXX, No. XXXX Suppl, Sxx–Sxx (2010). To cite previously published articles from the collection, please use the original citation, which can be found at the start of each article.

### VISIT THE OUTLOOK ONLINE

The *Nature Outlook Cancer Prevention* supplement can be found at <http://www.nature.com/nature/outlook/cancerprevention>

All featured articles will be freely available for 6 months.

### SUBSCRIPTIONS AND CUSTOMER SERVICES

For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242. Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea: Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

### CUSTOMER SERVICES

Feedback@nature.com  
Copyright © 2010 Nature Publishing Group



## INTRODUCTION

# The prevention agenda

*Despite our relative wealth of knowledge about the causes of cancer, the disease persists — and the burden is worsening. Prevention demands political will, ample funding and a change in mindset.*

BY TIFFANY O'CALLAGHAN

Half a century ago, more than 75% of British men smoked; today, that figure is closer to 20%. This drop has cut lung cancer deaths in middle aged men in the United Kingdom by as much as half. A similar trend, albeit less steep, is evident in other countries where smoking has declined, including the United States

And that's not the only significant advance in preventing cancer. Screening has also made an impact. The pap smear, to detect precancerous cells in the cervix, has helped cut US cervical cancer mortality rates from 5.5 per 100,000 in 1975 to just 2.4 in 2007. Antiviral vaccines are another success story: introduction of the hepatitis B virus (HBV) vaccine in 1982 cut chronic HBV infection rates among children in some countries from 15% to less than 1%, which has translated into reduced rates of liver cancer in adults. Hopes are high that vaccines against the human papillomavirus (HPV) will make similar inroads.

In spite of these encouraging trends, cancer — in all its guises — continues to undermine global health. In 2008, there were 12.7 million new cancer cases and 7.6 million deaths according to the American Cancer Society (ACS), costing the global economy nearly US\$900 billion. By 2030, the World Health Organization (WHO) predicts we'll face more than 21 million new cancer cases and 13 million deaths each year at skyrocketing costs to society.

The vast bulk of cancer research is trying to find treatments for people who are already sick. But approximately a third of cancers are caused by tobacco and at least a quarter are attributed to other lifestyle factors. The focus on cures only perpetuates the Sisyphean task of keeping cancer at bay. With all that we know about preventable causes of cancer, why is the incidence of cancer increasing? The answers are as complex and intertwined as the causes themselves.

## WILLPOWER REQUIRED

The sources of cancer are manifold. It's clear that there are environmental effects: second generation immigrants exhibit the disease patterns of their compatriots not of their ancestors<sup>1</sup>. But there's more to it than exposure. Although up to 90% of lung cancer is caused

by smoking, fewer than a sixth of smokers develop the disease. Some cancers are caused by faulty genes, such as *BRCA1/2* in breast and ovarian cancer; other cancers are fuelled by hormones such as oestrogen, testosterone or insulin. And then there are pathogens: the WHO estimates that 6% of cancers in wealthy nations and 22% in low- and middle-income countries are caused by viruses such as HBV, HPV and hepatitis C virus (HCV), bacteria such as *Helicobacter pylori* and waterborne parasites. Lifestyle affects cancer risk, too. In the past two decades as waistlines have expanded, so has the evidence linking obesity with the risk of breast, endometrial, colorectal and other cancers.

Yet even when the causes are understood, it is not easy to translate that knowledge into preventive actions. It was the 1950s when British epidemiologist Richard Doll proved smoking causes lung cancer, but it took decades to whittle away at the cigarette culture. Smoking restrictions imposed in the past 10 years were the result of a slow, incremental gathering of medical data and political will. Eventually "it became convincing to the public, at which point it was much easier to regulate," says David Hunter, an epidemiologist at the Harvard School of Public Health.

Smoking, however, is something of a special case. Building the political will for smoking bans in workplaces and public

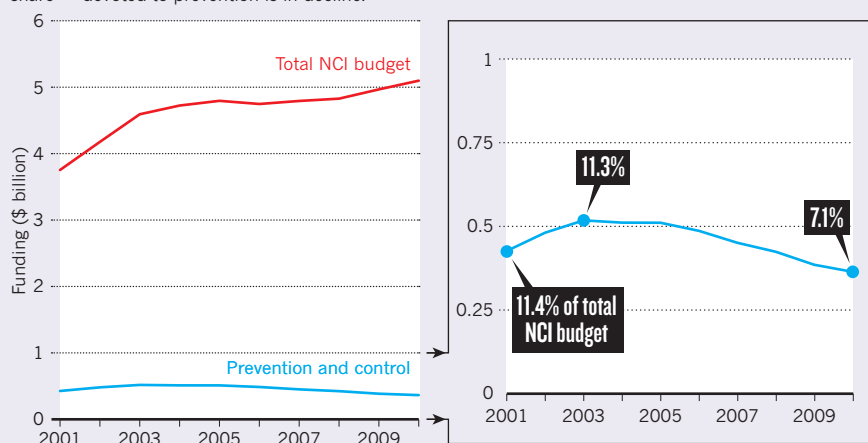
facilities hinged on the health risk of second-hand smoke to non-smokers. Could similar regulations be introduced when "there is not a direct cause and effect between your behaviour and my potential illness?" asks Hunter. Public health advocates can argue, for example, that obesity-related illnesses increase overall healthcare costs, but the logical steps from eating habits to obesity and then cancer risk are less straightforward. "That's a much more indirect case and it's harder to make," Hunter points out.

Future cancer prevention strategies might curtail individual choice — be it mandating vaccinations, banning trans-fats or taxing unhealthy food. These are treacherous political waters. "They are hard decisions that will not be popular," says Arnie Purushotham, an oncologist at the Integrated Cancer Centre (ICC), King's College London.

The unpopularity of such policies is evident from recent examples. Despite widespread efforts to promote access to the HPV vaccine, the ACS estimates that fewer than one in four girls who begins the course of vaccinations actually finishes it — partly owing to the social stigma associated with a cancer caused by a sexually transmitted infection. Efforts like the 2006 trans-fats ban in New York City are often decried as 'nanny state' meddling. Attempts to pass a 1% tax on sugary drinks in New York were ridiculed — in January 2011 the state's health commissioner

## PREVENTION'S DECLINING SHARE

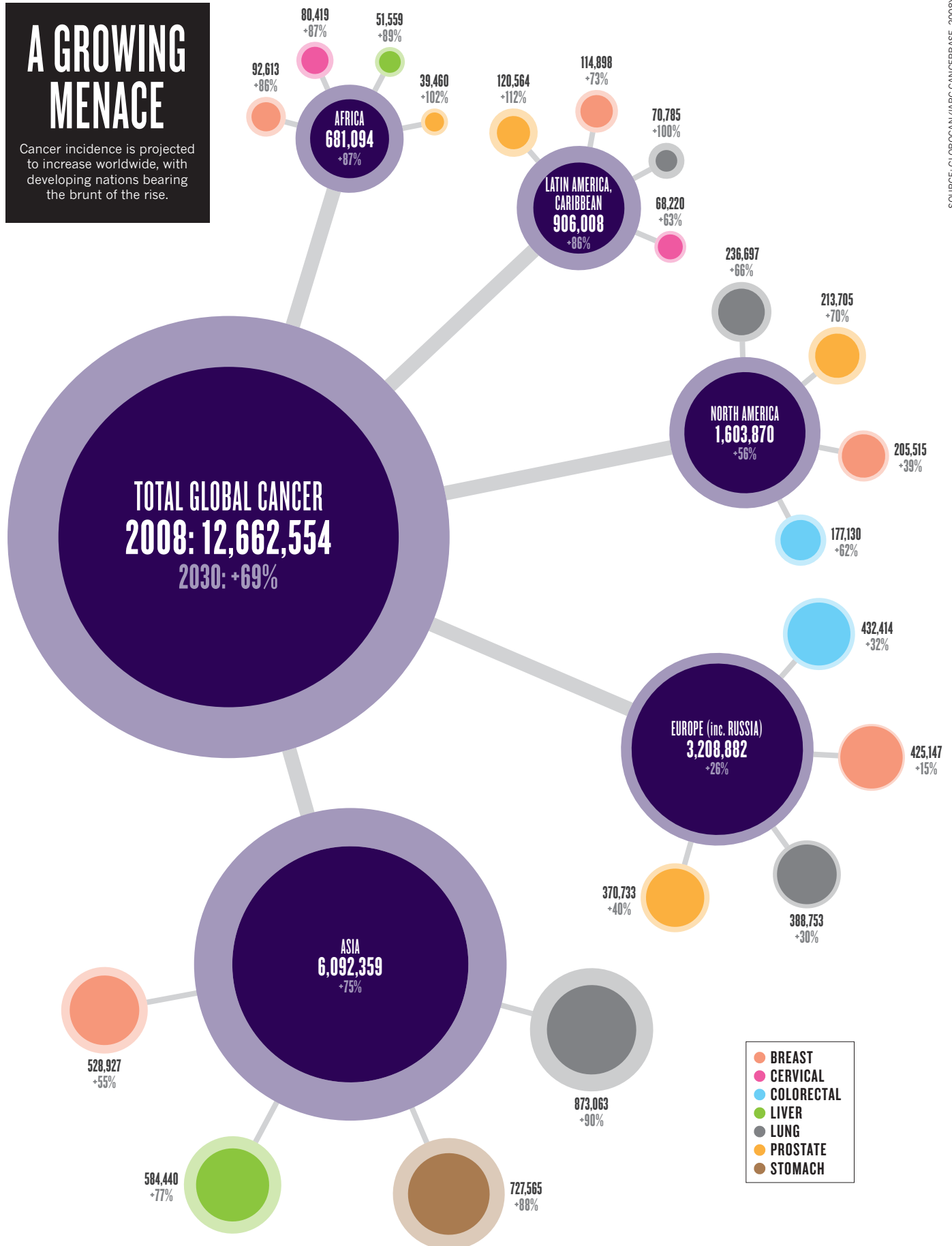
Although the budget of the US National Cancer Institute is steadily increasing, the amount — and the share — devoted to prevention is in decline.



SOURCE: NCI

# A GROWING MENACE

Cancer incidence is projected to increase worldwide, with developing nations bearing the brunt of the rise.



SOURCE: GLOBOCAN (IARC CANCERBASE, 2008).

said the tax was off the table, for now at least.

The broader notion of cancer as a preventable disease is not yet fully accepted. A study published in May 2010 found that the more local news coverage of cancer watched by Americans, the more likely they were to have fatalistic views of the disease<sup>2</sup>. Such ideas are even more problematic in a world of growing cancer risks, as developed countries export their bad habits. "Tobacco companies moved away from rich countries to poor countries," says Peter Boyle, head of the International Prevention Research Institute in Lyon, France. Higher rates of smoking, obesity and alcohol consumption mean low-income countries will struggle, but without the infrastructure to cope. "The poor countries of the world are going to be absolutely hammered in the next couple of decades by the diseases that are common in the developed countries now," says Boyle. And cancer is leading the charge.

### ELUSIVE FUNDING

Prevention research costs money, but funding decisions tend to be skewed in favour of developing treatments. Prioritizing prevention requires long-term thinking, yet government research goals can shift with each election. Ten years ago, 11.4% of the National Cancer Institute's annual budget was specifically allocated to cancer prevention and control. Since then, that allocation has steadily declined (see Prevention's declining share, page S2). According to a report by Purushotham and colleague Richard Sullivan for the G20 summit in 2010, less than 4% of worldwide public funding for cancer research goes to prevention.

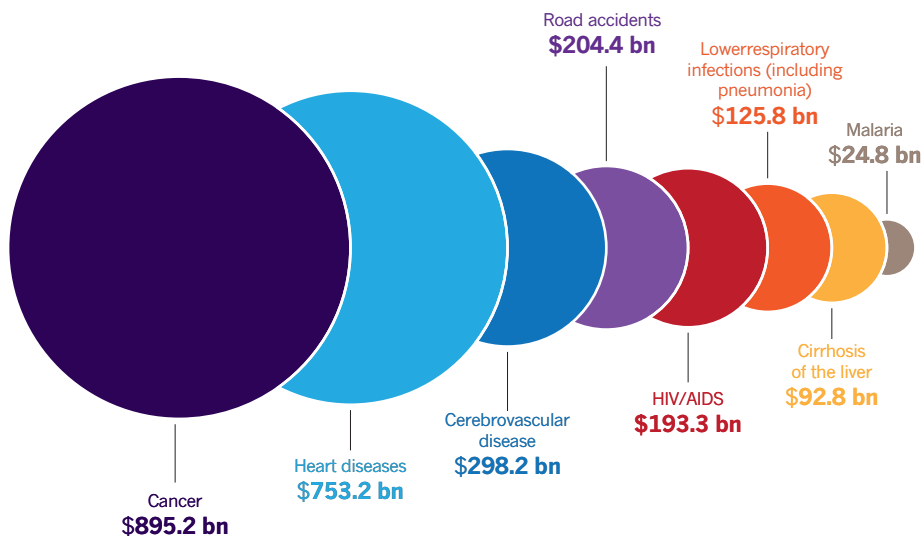
Some of the reasons for this distribution are obvious. The need for treatment is urgent, and survivors often go on to champion the cause. Prevention lacks this powerful advocacy group. "There is not a grateful patient to pressure the politicians to increase funding for the disease they have or have been cured from," says Hunter. In prevention, he adds, "the successes are virtual. With a cure, or even just a short increase in life expectancy, it feels more real for the public".

Moreover, prevention research "entails a very different form of research than setting up a lab and getting some mice and putting some carcinogen on them," says Ian Magrath, at the International Network for Cancer Research and Treatment, a not-for-profit organization based in Brussels, Belgium. "The type of research you have to do takes a much more fuzzy form because it involves human behaviour and psychology."

Industry financing models struggle too. Developing a new treatment can cost as much as US\$1.3 billion. That figure could be greater for medicines that have a higher safety requirement. When you take a preventive drug, "how

### COUNTING THE COST OF CANCER

The burden of cancer, calculated as the cost of years lost from ill-health, disability or early death, outweighs all other health concerns.



SOURCE: THE GLOBAL ECONOMIC COST OF CANCER (ACS, 2010).

much risk are you willing to take? Most people would say zero," says Kenneth Kaitin, a pharmacologist at Tufts Center for the Study of Drug Development in Boston, Massachusetts. Proving prevention is also harder and more time-consuming than proving treatment. "You have to have a large enough sample size of people who would eventually have cancer to

**"The poor countries of the world are going to be hammered in the next couple of decades by the diseases that are common in the developed countries now."**

prove that this is not just by chance," says Kaitin. "In essence you're proving the negative." Timing has other ramifications. Drug patents are filed in the early stages of clinical testing, and apply for 20 years or so, depending on extensions. After approval, which can take 10 years or more, a company has limited time before the drug goes generic. Long trials will eat into that time, reducing ability to recoup investment. And that is assuming they can negotiate reimbursement from insurers. Compared with drugs for treatment, Kaitin says, "it's even more cumbersome and onerous for a company to try to justify reimbursement to prevent a disease."

### LOOKING FORWARD

Despite these hurdles, cancer prevention advocates are pushing ahead. "You have to try to do something that is achievable," says Purushotham. Many research organizations are starting to infuse a prevention ethos into their medical approach. At the ICC, prevention messages are being added to the patient consultation process. If a patient comes in with a lump that turns out to be benign, for

example, Purushotham's team asks the patient about lifestyle factors that could elevate their cancer risk. The ICC also has a pilot programme underway for 'speed dating' between family doctors and cancer specialists, enabling oncologists to inform physicians about the latest cancer detection and prevention research.

In certain institutions, prevention is slowly moving up the list of cancer priorities. In 2009, of all the programmes at the Yale Cancer Center (YCC), the prevention programme received the largest slice of NCI funding. And to promote better collaboration between disciplines, researchers in the YCC have monthly meetings with colleagues they wouldn't otherwise meet — psychologists working with molecular geneticists, epidemiologists collaborating with clinicians — to share data and talk about new strategies. "We try to increase cross-talk," says Yong Zhu, co-director of Yale's cancer prevention programme. "We need to increase our communication among different research groups."

So the message is slowly being heard. "A prevention agenda is critical to have any kind of impact on the disease for the future," says Purushotham. As the global health community lays the foundation for future policies, perhaps gathering the necessary will for widespread cancer prevention is a matter of reminding ourselves of some age-old wisdom. As Thomas More, the sixteenth century philosopher, once wrote: "It is a wise man's part, rather to avoid sickness than to wish for medicines." ■

**Tiffany O'Callaghan** is a freelance writer based in London.

1. Boyle, P, Levin, B. (eds) *World Cancer Report 2008*. (IARC Press, 2008).
2. Niederdeppe, J. et al. *Journal of Communication* **60**, 230–253 (2010).

➔ **NATURE.COM**  
to read the latest  
news and research  
on cancer genomics  
[go.nature.com/5Gv7NO](http://go.nature.com/5Gv7NO)





## CHEMOPREVENTION

# First line of defence

*Combinations of drugs are showing some promise as therapeutic agents that stop cancer before it starts.*

BY LAUREN GRAVITZ

The mice in Xiangwei Wu's lab at the MD Anderson Cancer Center in Houston seem to be cheating death. Despite being genetically pre-programmed for colon cancer, they're staving off disease with a novel course of 'cancer cleansing'. Every three weeks, Wu injects the mice with a drug combo that targets only the mutant cells, prompting the cells to self-destruct. There are no side effects, nor is there any visible damage. These mice are receiving the most cutting-edge therapy conceivable: short-term treatment for long-term prevention.

Taking drugs to prevent cancer — rather than treat it — is not a new idea. In fact, over the past few decades a new specialty known as 'chemoprevention' has grown up around the concept. It's a tantalizing notion: swallow a pill that can thwart disease before it manifests. The reality, of course, may never be that simple.

An untold number of genetic changes can trigger cells to become cancerous — predicting what those changes will be and who will get them is still incredibly difficult (see Portents of malignancy, page S19). Studying the changes as they occur is even harder, at least in humans. Researchers cannot investigate a developing disease until they know it's there, which means that stopping it is a biomolecular game of Whac-a-Mole. Complicating the problem is the thorny issue of people who feel fine starting a long-term drug regimen — one that can cause troublesome side effects or even put them at risk for another disease.

The process of carcinogenesis can be exceedingly slow. "Cancer begins 20 years before a woman feels the lump in her breast," says Michael Sporn, a pharmacologist at Dartmouth and the person who introduced the concept of

cancer chemoprevention 35 years ago. That long span of time offers several opportunities for intervention. There's the mutation of a cell into something not-quite-healthy but also not-quite-cancerous. There's the transformation of a population of these precancerous cells into something that is immortal and grows unchecked. And then in most cancers, there's the metastatic movement of some of these cells away from the tumour, which establish a foothold elsewhere in the body.

A chemoprevention agent that blocks the very first step would be best. But most researchers would consider a drug successful if it could stop the disease progressing from any stage to the next.

## QUENCHING INFLAMMATION

Breaking the first link in the carcinogenesis chain is the ideal place to start: kill off any mutant cells that the immune system misses and do so before they have a chance to become cancerous. Yet this is a difficult proposition. Not only is it nearly impossible to observe these early changes in humans, but a drug for such early prevention would have to be absolutely benign in order to justify a near lifelong prescription.

Many in chemoprevention are beginning to think that perhaps the best way to catch cancer is to target inflammation. Chronic inflammation appears to encourage tumours by prompting the growth of new blood vessels and a remodelling of the extracellular matrix — creating a prime setting for normal cell growth to turn malignant.

Inflammation may be at the root of a host of serious ailments, from heart disease to diabetes. Stemming the tide of inflammation might prevent not only cancer but a number of other diseases. This theory led chemoprevention researchers to turn to two drugs — celecoxib and aspirin — that target the cyclooxygenase enzymes (COX-1 and COX-2), which play key roles in inflammation and pain. Both of the drugs are non-steroidal anti-inflammatories (NSAIDs), which epidemiological studies suggest may reduce the risk of colon and other cancers.

Aspirin inhibits COX-1, while celecoxib (Celebrex) inhibits COX-2. COX-1 is produced in tissues throughout the body, and is known to mediate the production of prostaglandins — chemical messengers that control a number of physiological functions, such as lowering blood pressure, regulating body temperature and controlling inflammation. COX-2, on the other hand, is strictly regulated and tends to spike during inflammation and other stress — an abundance of COX-2 has been linked to the growth and proliferation of cancerous and pre-cancerous cells. Inhibiting the COX pathways can alter cancerous and precancerous cells by decreasing blood vessel formation and cell growth. COX inhibition enhances a mutant cell's ability to commit

➔ **NATURE.COM**  
to find the latest  
research papers on  
novel cancer drugs  
[go.nature.com/mk1JKp](http://go.nature.com/mk1JKp)

suicide in a process known as apoptosis and enables the immune system to recognize and target the cells for destruction.

Celecoxib — along with other COX-2 inhibitors such as rofecoxib (Vioxx), since removed from the market — were developed as alternatives to aspirin, which can cause gastrointestinal bleeding after prolonged use. And, for a while, they seemed promising not just for pain but also for cancer prevention. The US Food and Drug Administration (FDA) approved daily doses of celecoxib for reducing colon cancer risk in people with a rare genetic disease called familial adenomatous polyposis (FAP), which causes dense outcroppings of intestinal polyps. Early reports indicated that daily doses of COX-2 inhibitors could decrease the risk of breast, skin and colon cancer in high-risk individuals. While COX-2 inhibitors are still a promising drug for chemoprevention, no drugs that work through this pathway are now on the market.

But it became apparent that prolonged use of COX-2 inhibitors increased the risk of stroke and heart attack, so the FDA banned most from the market. Celecoxib remained available until early February 2011, when the manufacturer voluntarily removed labeling for this use due to its inability to do follow-up studies. A number of cancer prevention researchers continue to pursue COX-2 inhibitors, trying to develop compounds with the same inflammation- and cancer-fighting effects as celecoxib but without the risk of cardiovascular problems.

Aspirin comes with its own set of risks. “People die from it,” says Leslie Ford, associate director for clinical research in the cancer prevention division of the National Cancer Institute (NCI) in Washington. “There are about 16,000 deaths a year from gastro-intestinal bleeds in people taking aspirin. But these risks

aren’t preventing clinical trials, and the evidence is mounting that aspirin can decrease a person’s risk of colon cancer, as well as lung, prostate and even brain cancer.

Other systemic, inflammation-targeted drugs may also help ward off a variety of cancers. Some studies suggest that statins — developed for cholesterol management — might disrupt the growth and proliferation of cancer cells. Reports suggest that medications taken to increase insulin sensitivity in type 2 diabetes could lower risk for several types of cancers, including head, lung and neck. Such drugs include metformin, pioglitazone and rosiglitazone (although the latter has recently been linked to heart problems).

Precisely how or why these drugs might aid cancer prevention remains a mystery. “Every time I think I have a specific agent for a specific pathway, when it’s tested, I find all sorts of other activities that may play into it as well,” says Ernest Hawk, head of the Division of Cancer Prevention & Population Sciences at MD Anderson Cancer Center. “This really gets at the question: do we really ever know the mechanism of anything?”

### CURBING PREMALIGNANT CELLS

Currently, most chemoprevention aims to prevent premalignant cells from completing the process of carcinogenesis. These agents have clear cellular targets and are intended to treat people at high risk — those with a family history of disease or a known genetic mutation, or who are already known to have precancerous cells.

Because different cancers evolve in various ways, a one-size-fits-all approach is not on the horizon. “One of the biggest challenges of this field is that it’s really not one field,” says Eva Szabo, a researcher in the cancer prevention division of the NCI. “Before we understood the

complexity that is cancer, the thought was that we could have a generalized strategy that could prevent the transformation of cells, or could arrest progression.” Now, she says, it’s apparent that lung or breast or colon cancers can take multiple forms. And because the pathogenesis is different, the prevention has to be different, too.

Most of the medications approved by the FDA for the purpose of treating precancerous cells or reducing cancer risk (see Table opposite) fall into this category. Women who fall into high-risk categories for breast cancer, for example, can take raloxifene (Evista) or tamoxifen (Nolvadex), both of which have been shown to cut a woman’s risk of estrogen-receptor positive breast cancers by as much as 50%. And five different creams and ointments have been approved to prevent skin lesions from developing into skin cancers such as squamous cell carcinoma.

Xiangwei Wu’s colon cancer-prone mice belong in this category, too. Wu’s two-drug combination takes advantage of mutations specific to precancerous cells in the colon, prompting only those mutant cells to self-destruct. “Because most preventive drugs don’t get rid of the bad cells, people have to be on them continuously for a long time,” says Wu. Such is the case with raloxifene and tamoxifen.

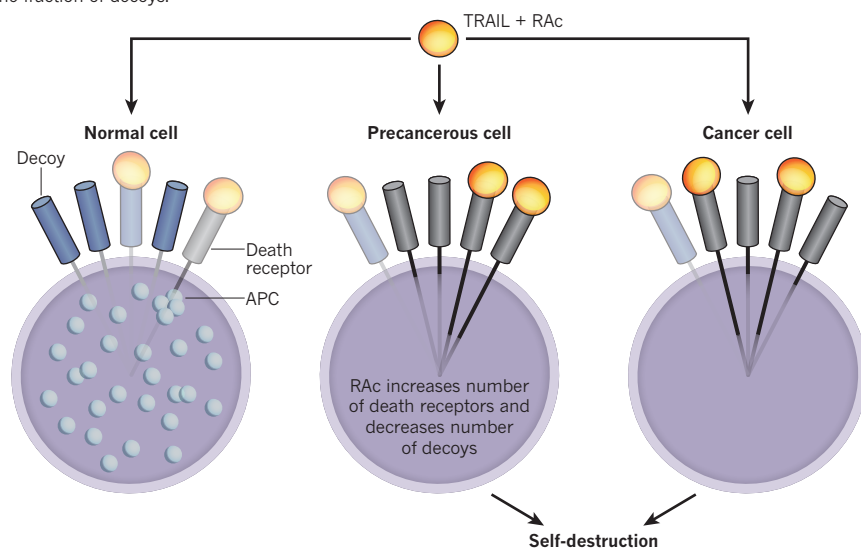
Wu’s mice seem to be experiencing the best of all possible outcomes: they receive intermittent therapy that kills emerging cancer cells, but benefit from breaks between doses to recover from any side effects. In addition, the two agents — tumour necrosis factor-related apoptosis-inducing ligand (TRAIL) and all-trans-retinyl acetate (Rac) — work synergistically. Most researchers see this as important in preventing a rebound of drug-resistant populations. TRAIL has already proven promising for treating cancer because it appears to leave healthy cells unharmed while inducing cancer cells to self-destruct. In combination with Rac, Wu has found, TRAIL can induce precancerous cells to commit suicide, too (see Cancer-stopping combo).

Another two-drug combo aimed at colon cancer is one of the most promising chemoprevention prospects in the pipeline. Low daily doses of sulindac, an NSAID, along with difluoromethylornithine (DFMO) appear to do far more good together than either does alone — and with minimal toxicity, according to researchers at the University of California, Irvine. “We wanted to find the lowest dose of each at which we could find a relevant effect in the colon,” says Frank Meyskens, director of the university’s Chao Family Comprehensive Cancer Center.

The combination is one of push and pull. Cancerous cells have a hard time regulating the metabolism of polyamines and so have abnormally high levels of these compounds, which healthy cells use for growth and development. DFMO prevents polyamine synthesis, while

### CANCER-STOPPING COMBO

TRAIL binds to receptors to trigger cells to self-destruct. Normal cells are protected both by decoy receptors and by the protein APC, which blocks the death-signal pathway. Rac helps kill precancerous cells by lowering the fraction of decoys.





## FDA-APPROVED CHEMOPREVENTION DRUGS

| Drug  | Brand name(s)                  | Cancer type                    | Year first approved | Target / mechanism   | Dosing (how long, how often)   | Original manufacturer         |
|---|--------------------------------|--------------------------------|---------------------|--|--|-------------------------------|
| Tamoxifen   | Nolvadex<br>Istubal<br>Valodex | Breast                         | 1998                | Selective estrogen receptor modulator (SERM)   | Daily, for 5 years   | AstraZeneca                   |
| Raloxifene  | Evista                         | Breast                         | 2007                | SERM   | Daily, duration unknown  | Eli Lilly                     |
| HPV vaccine   | Gardasil<br>Cervarix           | Cervix<br>Vulva<br>Vagina/Anus | 2006                | Elicit immune response to prevent infection by the most common cancer-causing types of HPV                                       | 3 doses over the course of 6 months  | Merck & Co<br>GlaxoSmithKline |
| Porfimer sodium + photodynamic therapy (PDT) & omeprazole | Photofrin                      | Esophageal                     | 2003                | Lodges in precancerous cells and upon exposure to certain light produces an active form of oxygen that kills nearby cancer cells | Single injection, followed by light therapy three days later.<br>Can be repeated after 90 days               | Axcan                         |
| Fluorouracil  | Efudex<br>Fluoroplex<br>Carac  | Skin                           | 1970                | Interferes with DNA synthesis and leads to cell death  | Apply to affected areas twice daily until lesions are gone, as long as 10-12 weeks.                          | Valeant                       |
| Diclofenac sodium 3%                                      | Solaraze                       | Skin                           | 2000                | Exact mechanism is unknown   | Apply to lesion twice daily, for 60-90 days  | PharmaDerm                    |
| 5-aminolevulinic acid + PDT*                              | Levulan                        | Skin                           | 1999                | Solution kills precancerous cell when exposed to light <sup>1</sup>  | Apply a topical solution to lesion, then single photodynamic therapy treatment 14 to 18 hours later.         | DUSA                          |
| Imiquimod   | Aldara (5%)<br>Zyclara (3.75%) | Skin                           | 2004                | Enhances immune response and promotes apoptosis  | Zyclara: Applied topically for two 2 weeks, then a 2 week break<br>Aldara: Applied twice a week for 16 weeks | Graceway Pharmaceuticals      |

sulindac triggers cells to purge it. Together, in low doses, the two compounds work to lower polyamines in precancerous tissue; the result is fewer precancerous growths, which are risk factors for colon cancer. In clinical trials, Meyskens' team showed that after three years of therapy, the combined treatment reduced the occurrence of polyps by 70% and a greater than 90% reduction in advanced adenomas—the ones most likely to go on to develop cancer. The researchers found a company, Cancer Prevention Pharmaceuticals, to take the treatment through phase III trials into the market.

## PREVENTING METASTASIS

With most solid-tumour cancers, the biggest danger is not the tumour itself but its ability to metastasize. While the primary tumour continues to grow, rogue cells break off, work their way into the blood stream, and move on to colonize other areas of the body. Metastases represent an advanced accumulation of genetic damage, which are too many mutations for conventional drugs to target at once.

"Across all cancers, over time, there is an almost exponential increase in the number of genetic mutations," says Raymond Bergan, a specialist in preventive oncology at Northwestern University medical school in Chicago. "We do a pretty good job of making a drug that can hit a target. But making one drug that will hit two different targets is very difficult."

Bergan is aiming for a target that prevents growth and metastasis of a primary tumour. His lab has been investigating genistein, a soy isoflavone that has been sold as a nutritional supplement for years. Bergan's group is putting the compound through its paces in the lab, and

it is showing promise in trials for preventing — even reversing — the metastatic process of prostate cancer.

For cancer to metastasize, tumour cells must detach from their neighbours. This decreased adhesion is at least partially mediated by the enzyme focal adhesion kinase. Studies show that genistein blocks the activation of this enzyme so that prostate cancer cells remain tethered to the tumour.

Genistien also prevents mobile prostate cancer cells from invading tissue. As healthy cells grow and divide, the enzyme matrix metalloproteinase-2 (MMP-2) helps break down extracellular membrane proteins to make way for new growth. One of the proteins it degrades, however, is first in line for attack by invading cancer cells. Researchers have found that higher concentrations of MMP-2 correlate with poor prognosis. Genistein counteracts this by targeting a protein that increases MMP-2 concentrations. By binding to and inhibiting this protein, it can prevent production of MMP-2.

In phase II clinical trials, Bergan has shown that genistein can decrease MMP-2 in human prostate tissues — he's now looking at whether this prevents cancer cells from moving beyond the prostate, thereby reversing the cancer's evolution into metastatic disease. Following up on Bergan's research, Seema Khan, also at Northwestern University, is studying whether genistein — as part of a mixture of soy isoflavones — might prevent the spread of breast cancer. Her results, she says, are far less conclusive and hint that in younger women, this treatment could lead to a slight increase in cell proliferation.

Indeed, uncertainty and ambiguity are the norm in cancer prevention research. "I believe that there are enough instances where chemoprevention has been deleterious that it calls into question how much we understand," says John Potter, a senior advisor at the Fred Hutchinson Cancer Research Center in Seattle. "If I were to write a prescription for the field, I'd want to match risk and benefit."

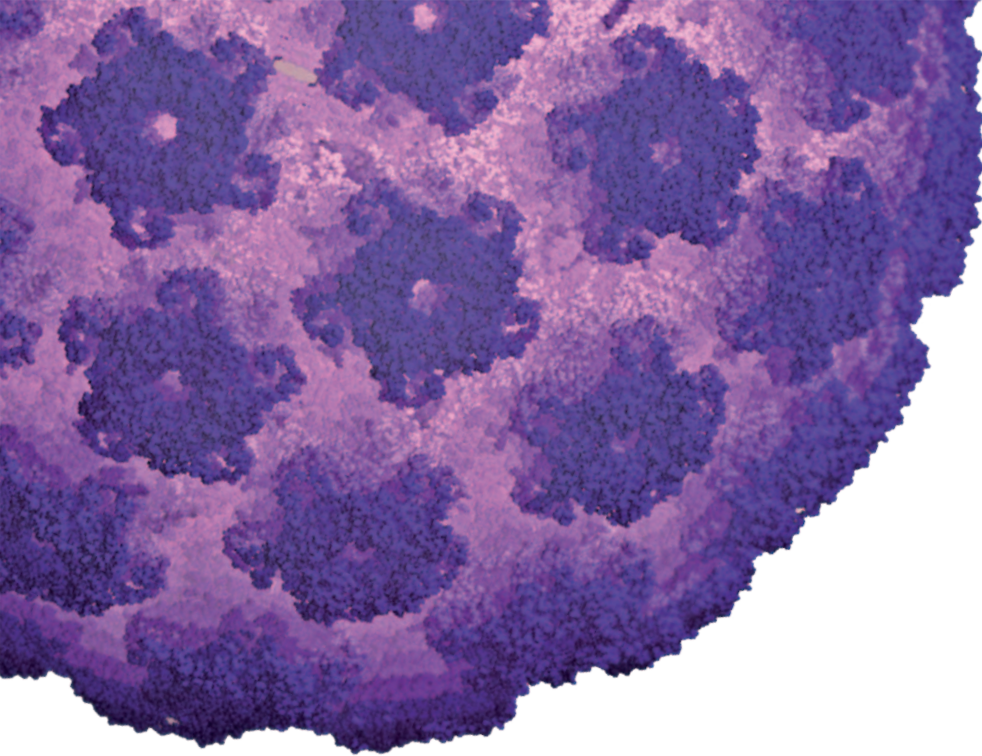
Tailoring the right therapy to the right risk profile will help. "Most people think of cancer prevention like preventing polio: you get a one-time shot and you never have to worry about it again," says Powel Brown, chair of Clinical Cancer Prevention at MD Anderson. "But it's more likely to be akin to taking anti-cholesterol medicine for the prevention of heart disease." That is, you'll be taking the medication indefinitely — and while there may be side effects, they will be acceptable to reduce the risk of a potentially fatal disease.

Indeed, if scientists and clinicians are able to establish a cancer prevention system as well-tuned as that for heart disease, it would count as a massive success. For now, researchers in the field will have to look to the mice in Wu's lab, which are revealing one possible path to a cancer-controlled — if not cancer-free — future. ■

**Lauren Gravitz** is a writer in Los Angeles.

1. Bode, A. M. *et al. Nature Reviews Cancer* **9**, 508–516 (2009).
2. Lippmann, S. M. *et al. Cancer Res.* **69**, 5269–5284 (2009).
3. Rothwell, P. M. *et al. Lancet* **377**, 31–41 (2011).
4. Zhang, L. *et al. Nature* **464**, 1058–1063 (2010).
5. Germer, E. *et al. J. Clin. Oncol.* **15**, 758–761 (2009).
6. Xu, L., *et al. J. Natl. Cancer Inst.* **101**, 1141–1155 (2009).





## VACCINES

# Know your enemy

*Vaccines are arguably our greatest medical achievement. But to what extent can they help prevent cancer?*

BY MICHAEL EISENSTEIN

Cancer operates like a well-disguised saboteur. Occasionally it slips up by displaying unusual proteins, tripping immunological surveillance systems that are checking for abnormal growth. For decades now, scientists have tried to exploit this vulnerability with therapeutic vaccines — injections of tumour-associated proteins that essentially hang a ‘Wanted’ poster, helping immune cells recognize and kill cancer cells.

After a string of expensive and dispiriting defeats, therapeutic cancer vaccines recently registered their first big win. In April 2010, the US Food and Drug Administration (FDA) approved Provenge (sipuleucel-T) — a mixture of a patient’s own cells incubated with a protein expressed by 95% of prostate tumours. This was not an unequivocal victory, however. “On average, patients live about four months longer,” says Martin Kast, a cancer vaccine specialist at the Norris Comprehensive Cancer Center at the University of Southern California (USC) in Los Angeles. “It certainly measures up to many chemotherapeutics, but there’s still a long way to go.”

The problem is that as cancer develops, it disrupts the immune system to prevent an effective counter-attack. Accordingly, Kast and other researchers have been refocusing their efforts towards developing cancer vaccines as

preventative measures, before cancer sets its traps. “You don’t wait to get polio before taking the polio vaccine — you vaccinate prior to the engagement of the pathogen,” says Vincent Tuohy, an immunologist at the Cleveland Clinic in Ohio. With therapeutic vaccines “we were asking the immune system to go in and heroically eliminate this large and mature tumour load, and that’s a problem. If you want to get rid of a disease, you do it prophylactically”.

## CATCHING CANCER

So far, the best examples of anticancer vaccines are those that thwart cancer-causing infections. For example, vaccination against the hepatitis B virus (HBV) offers lasting protection against liver cancer: an estimated 54% of hepatocellular carcinoma (HCC) cases are attributable to HBV. A 20-year study in Taiwan demonstrated that vaccination reduces the risk of developing HCC by roughly 70% (ref. 1).

Vaccines against human papillomavirus (HPV, see main image) can potentially have even greater impact. HPV is responsible for almost all of the half-a-million cases of cervical cancer worldwide, as well as 60,000 cases of anal, genital and throat cancer. The two FDA-approved vaccines, Gardasil (Merck,

approved 2006) and Cervarix (GlaxoSmith-Kline, approved 2009), both demonstrate remarkable efficacy in preventing infection by HPV strains 16 and 18, which account for the lion’s share of HPV-associated cancer. “I’m optimistic that we’re going to have long-term protection,” says John Schiller, head of the Neoplastic Disease Section at the National Cancer Institute in Bethesda, Maryland, and a co-inventor of both vaccines.

At least a dozen other strains of carcinogenic HPV elude these vaccines, however. The reach of a vaccine is determined by its valency, which is the number of different targets it presents to the immune system. Gardasil is effective against four strains of HPV, Cervarix two. Merck is looking to up the ante: its V503 vaccine, now in a phase III clinical trial, covers nine strains. Meanwhile, Schiller is developing an ‘omnivalent’ vaccine, derived from a different viral protein, which could offer complete protection. In tissue cultures, Schiller says, this vaccine “prevents infection by all the HPV types we’ve ever tried”. Schiller anticipates that such a broad antiviral vaccine could be routinely given to younger patients of both sexes, greatly expanding coverage — as of 2008, only 18% of girls aged 13 to 17 years in the United States had received a full course of HPV vaccinations. Sanofi-Pasteur is likely to advance such an omnivalent approach into clinical trials in the near future.

Scientists are also starting to make progress in developing vaccines against *Helicobacter pylori*, a bacterium linked to 60% of the one million or so cases of stomach cancer worldwide. In 2008, a team led by Peter Malfertheiner, a gastroenterologist at the Otto-von Guericke Universität in Magdeburg, Germany, showed that a candidate vaccine — now under development at Novartis — was both safe and capable of raising a strong immune response against selected proteins<sup>2</sup>. “These antigens are key players in the pathogenesis of *H. pylori* infection,” says Malfertheiner, “and we found a very significant systemic response.” His team plans to expose healthy volunteers to a moderately virulent strain of the bacteria to test the vaccine’s protective capacity.

## FRIEND OR FOE

Most cancers — including major killers such as lung and colorectal cancer — are not caused by infections. In these cases, prophylactic vaccines must target essential proteins that the tumour needs to thrive. Fortunately, the past decade has seen great progress in both cataloguing tumour proteins and developing genetically modified mouse models that closely mimic human cancer progression. Pier-Luigi Lollini, a molecular oncologist at the University of Bologna in Italy, reports early promising results from vaccinating mice against HER2/neu, a protein over-expressed in many breast tumours<sup>3</sup>. “We can prevent

➔ **NATURE.COM**  
for the latest  
research on HPV  
vaccines  
[go.nature.com/aaDgPM](http://go.nature.com/aaDgPM)

tumour onset, and the ongoing carcinogenic process fuelled by *HER2*, in mice for most of their adult life," says Lollini. His focus is purely on animal studies, but colleagues Guido Forni and Federica Cavallo are pursuing a parallel *HER2/neu* vaccine strategy in humans that could soon enter clinical trials.

Most cancer proteins are, to a certain extent, produced by healthy cells and as such enjoy a privileged 'self' status that protects them against immunologic attack. By incorporating carefully selected adjuvant molecules that enhance the immune response, vaccines can break this inherent tolerance and turn immunity against these self-proteins — without the nasty effects seen in autoimmune diseases. For example, a 2010 study by Tuohy and colleagues showed striking success in a mouse model of breast cancer by targeting the milk protein  $\alpha$ -lactalbumin<sup>4</sup>. This protein is normally expressed only during late pregnancy and lactation, but Tuohy notes that expression is also common in newly formed tumours. "One of the things they do is make inappropriate proteins like  $\alpha$ -lactalbumin," says Tuohy. In fact, vaccinated mice achieved 100% protection against breast cancer, provided they were dosed before tumours began to develop. Tissue damage and inflammation were limited to the breast tissue of nursing animals. This should not be a problem for humans, as the highest risk breast cancer patients are generally past childbearing age. "97% of women 'retire' their breasts from lactation after 40, and that's exactly the age range where 95% of breast cancers occur," says Tuohy.

Some researchers contend that broader protection could be possible with vaccines that target many different tumours. "You could potentially come up with multi-antigen approaches that cover 90% or even almost 100% of populations of different tumour types," says Mary Disis, head of the Tumor Vaccine Group at the University of Washington in Seattle. Immunologist Olivera Finn's group at the University of Pittsburgh has worked extensively with one such candidate, mucin-1

(MUC1). "About 80% or more of all human cancers aberrantly express this protein," says Finn. "The immune system sees that abnormal expression and generates a response." This response on its own is too weak to prevent cancer onset, but Finn and colleagues have developed a MUC1 vaccine that prepares the immune system in advance of tumour formation. Early trials have demonstrated a strong and potentially protective immune response in more than 50% of patients. Finn's group is conducting a clinical trial of MUC1 in Pittsburgh with patients at high risk for colorectal cancer — one of the few prophylactic cancer vaccine trials currently underway.

### TRIALS AND TRIBULATIONS

Prophylactic cancer vaccine trials face numerous obstacles, including the long-term endpoints needed to determine prevention in cancer-free individuals. Schiller's work with cervical cancer vaccines has benefited from the fact that lesions arising from HPV infection are a powerful predictor of cancer risk. Gardasil and Cervarix "haven't formally been demonstrated to prevent the cancer," he says, but preventing 98% of HPV-associated lesions represents a "good surrogate marker for protection".

Finn and other researchers are using similar pathological features to select individuals likely to acquire malignancies in the near term. Finn's MUC1 colon cancer prevention trials, for example, focus on patients with advanced adenomatous polyps — unusual lumps on the colon wall. Such growths, explains Finn, represent "the latest stage of premalignant changes, most proximal to colon cancer". Likewise, Kast's team at USC vaccinates mice shortly after they develop precancerous growths in their prostate<sup>5</sup>. This approach straddles the line between therapy and prevention: the vaccines flag existing abnormalities but nevertheless train the immune system to block the advance of full-blown cancer. Timing appears to be the crucial element that separates a failed therapeutic vaccine from a successful prophylactic

one. "If you vaccinate early on, before there is actual cancer but while carcinogenesis is underway, you have a chance," says Kast. Results from his studies in mice are promising. "Instead of dying between 6 and 12 months, most mice were still alive at 1.5 years, and they were dying of old age, not prostate cancer."

Even with a head start from preselecting high-risk patients, human studies will be a lengthy and expensive slog — a major deterrent for many companies and funding agencies. This is especially problematic for infections such as *H. pylori*, which primarily affects the developing world. "The calculation of the cost-benefit ratio is critical," says Malfertheiner. "The developing world does not have the money to buy vaccines, whereas the developed world has good therapeutics that can treat infection." Even existing vaccines such as Gardasil remain a costly proposition for poor nations, although improved formulation and low-cost generics may address this in the near future.

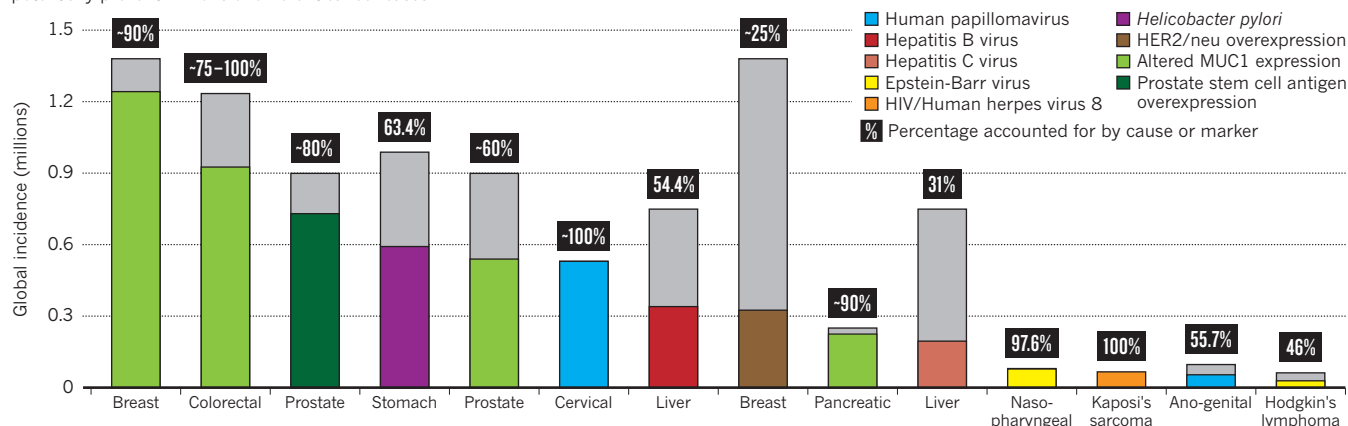
Exposure for cancer vaccines also remains a problem. The first dedicated conference, scheduled for March 2011 at Arizona State University in Tucson, was abruptly cancelled in February. "We were described as being too far ahead of the curve," says Kast. However, a symposium on preventative cancer vaccines in April 2011 at the annual meeting of the American Association for Cancer Research in Florida should help stimulate interest. For researchers, promising results from early stage multi-antigen studies speak for themselves. "When I first got into the field of tumour immunology 15 years ago, I remember saying, 'I don't think I'll ever see a vaccine to prevent cancer,'" recalls Disis. "I may have to eat my words." ■

Michael Eisenstein is a freelance journalist in Philadelphia.

1. Chang, M. H. *et al.* *J. Natl. Cancer Inst.* **101**, 1348–1355 (2009).
2. Malfertheiner, P. *et al.* *Gastroenterology* **135**, 787–795 (2008).
3. De Giovanni, C. *et al.* *Cancer Res.* **64**, 4001–4009 (2004).
4. Jaini, R. *et al.* *Nat. Med.* **16**, 799–803 (2010).
5. Gray, A. *et al.* *Vaccine* **27** Suppl. 6, G52–G59 (2009).

### VACCINE-PREVENTABLE CANCERS

Current vaccines against human papillomavirus and hepatitis B virus have already made an impact on the incidence of certain cancer types. Future vaccines can potentially prevent millions of different cancer cases.





## PERSPECTIVE

# The big C — for Chemoprevention

Drugs to prevent cancer are clearly possible despite some early missteps, says **Michael B. Sporn**. Restoring the cooperative ethos of decades past will help get us there.

IMAGES.COM/CORBIS



When a complex system starts to dysfunction, it is generally best to fix it early. The alternative often means delaying until the system has degenerated into a disorganized, chaotic mess — at which point it may be beyond repair. Unfortunately, the general approach to cancer has ignored such common sense. The vast majority of cancer research is devoted to finding cures, rather than finding new ways to prevent disease.

The results of these skewed priorities are plain to see: forty years after President Richard Nixon declared war on cancer, the death tolls from most common forms of cancer in the United States have not fallen. It's true that for some cancer types, mortality rates (adjusted for population size) have dropped during these decades, but there are huge, unhappy exceptions: mortality rates for lung and pancreatic cancer have stayed level since 1970, and the total number of US deaths each year from those diseases has doubled<sup>1</sup>.

Looking at these discouraging statistics, it is clear that something needs to change. We have been looking at the very nature of cancer in the wrong way. Breast cancer doesn't begin when a lump is first felt or detected by mammogram. All the common epithelial cancers (lung, colorectal, breast, prostate, pancreas and ovary), which account for the majority of deaths, have a long latency period — often 20 years or more. By the time they are clinically detectable, the cells in such carcinomas may harbour hundreds of mutations in different genes<sup>2</sup>. These cells provide no simple, single target for therapy. In contrast, during the long latency period, there is ample opportunity to use multi-functional, multi-targeted preventive drugs that block the development of invasive and metastatic disease.

That's the basic idea of cancer chemoprevention (see First line of defence, page S5): to arrest or reverse the progression of pre-malignant cells towards full malignancy, using physiological mechanisms that do not kill healthy cells. In experimental animals, it is now possible to prevent the onset of cancer in almost all the common organs in which human carcinoma occurs. Even more importantly, chemoprevention has now been validated in people. One class of drugs, known as selective estrogen receptor modulators (SERMs), can deliver as much as a five-fold reduction in incidence of estrogen receptor-positive breast cancer in women. These compounds — most notably tamoxifen, raloxifene and lasofoxifene — have the added benefit of suppressing osteoporosis<sup>3</sup>. Fenretinide, for which we have 15 years' worth of data, provides significant prevention of breast cancer in premenopausal women<sup>4</sup>. Two anti-androgenic agents, finasteride and



dutasteride, have been shown to be effective at reducing incidence of prostate cancer in long-term clinical trials<sup>5</sup>.

And yet we have a paradox: within the world of clinical oncology, chemoprevention of cancer is perceived to be a failure. As a result of some poorly designed and executed clinical trials over the past decade, scepticism abounds about the practicalities of chemoprevention. This harsh assessment is the conventional wisdom among groups as diverse as the pharmaceutical industry, the hospital establishment, the insurance industry, women's advocacy groups and the clinical oncology community itself. Of particular disappointment to many advocates of chemoprevention has been the general lack of enthusiasm from large pharmaceutical firms, as exemplified by the recent decisions of two major companies to curtail further development of lasofoxifene and arzoxifene, another highly promising SERM<sup>6</sup>. Many factors have contributed to this negativity, including difficult regulatory approvals, duration of patent protection and the omnipresent fear of liability in treating supposedly healthy people with drugs.

But attitudes toward chemoprevention need to be re-examined. Most fundamental is the bizarre misperception that people are 'healthy' until they have actual symptoms of invasive cancer, the corollary being that it is unwise and perhaps unethical to give a preventive drug to a healthy person. In reality, however, a person harbouring a premalignant lesion is not healthy, in spite of the absence of symptoms. Many of these people will go on to develop life-threatening cancers. The barn in which hay is smoldering before it bursts into flames is not a safe place.

Another canard is that cancer prevention efforts are not cost-effective. The argument is that the number of lives saved with a preventive drug would be too small with respect to the total number of people who need treatment. But this is a curious perspective. The number of houses destroyed by fire is trivial compared with the total number of houses, and yet almost every homeowner insures against fire. The conceptual problem here is that everyone doesn't die of cancer in a short period; this is a lifetime problem.

There is a simple answer: we should stop doing clinical chemoprevention trials in large populations of people at relatively low risk, and instead focus on cohorts at the highest risk. There are many such groups: women with *BRCA* mutations that can lead to breast and ovarian cancer, people with premalignant pancreatic lesions and

those with severe premalignant lung lesions (especially in current or former heavy smokers). Chemoprevention trials on such

groups will provide much more definitive results and with much less effort.

More broadly, the elevation of cancer prevention requires several actions. First, we need a massive educational effort to encourage the general public — not just special interest groups — to support prevention efforts. This has been very successful in cardiology; indeed, the pharmaceutical industry spends huge amounts of money on educational and advertising efforts to promote chemoprevention of cardiovascular disease with statins and anti-platelet agents. Unfortunately, these companies seem unwilling to similarly promote cancer chemoprevention.

In addition, we need to be vigilant about the safety of the preventive drug testing regimen. Thus, we should build in 'rest periods' in clinical chemoprevention trial design. Many drugs used for chemotherapy have severe toxicities if used long term, so rest periods are necessary and routinely used. Although, as a class, chemopreventive drugs are much less toxic, rest periods would make relatively safe drugs even safer. Moreover, drugs need to show extensive efficacy in animal experiments before undergoing human trial — wisdom that was forgotten in the failed clinical trials of beta-carotene, selenium and tocopherol.

Regulatory accommodations will also help. The Food and Drug Administration still forbids the use of two experimental drugs in clinical prevention trials, in spite of the fact that there is overwhelming evidence that combinations can be much safer and more effective than single agents<sup>7</sup>. Further appreciation and understanding of the concept of 'risk' is also essential. For new drugs, the proper comparison is not risk versus benefit but rather risk versus risk<sup>8</sup> — that is, the risk of doing nothing (which may have a deadly outcome), versus the risk of taking a preventive drug for long periods. Oncologists could follow the lead of cardiologists, who have developed a handy scorecard that numerically quantifies a patient's risk<sup>8</sup>.

On the basic science front, we must develop new multifunctional drugs that aim at entire networks in the body, rather than single targets<sup>9</sup>. We need further studies on the importance of epigenetics<sup>10</sup> and the tumour microenvironment<sup>8</sup> to develop chemopreventive drugs. The tumour microenvironment, with all of its stromal and inflammatory cells, is an essential part of a carcinoma. Major advances over the past decade in these areas are leading to the development of important new drugs for cancer prevention<sup>8,10</sup>.

So the challenges are numerous and daunting. There is great interest in personalized medicine (a wonderful goal) — but how can we do this successfully if the parameters we assess are exclusively genetic? Our

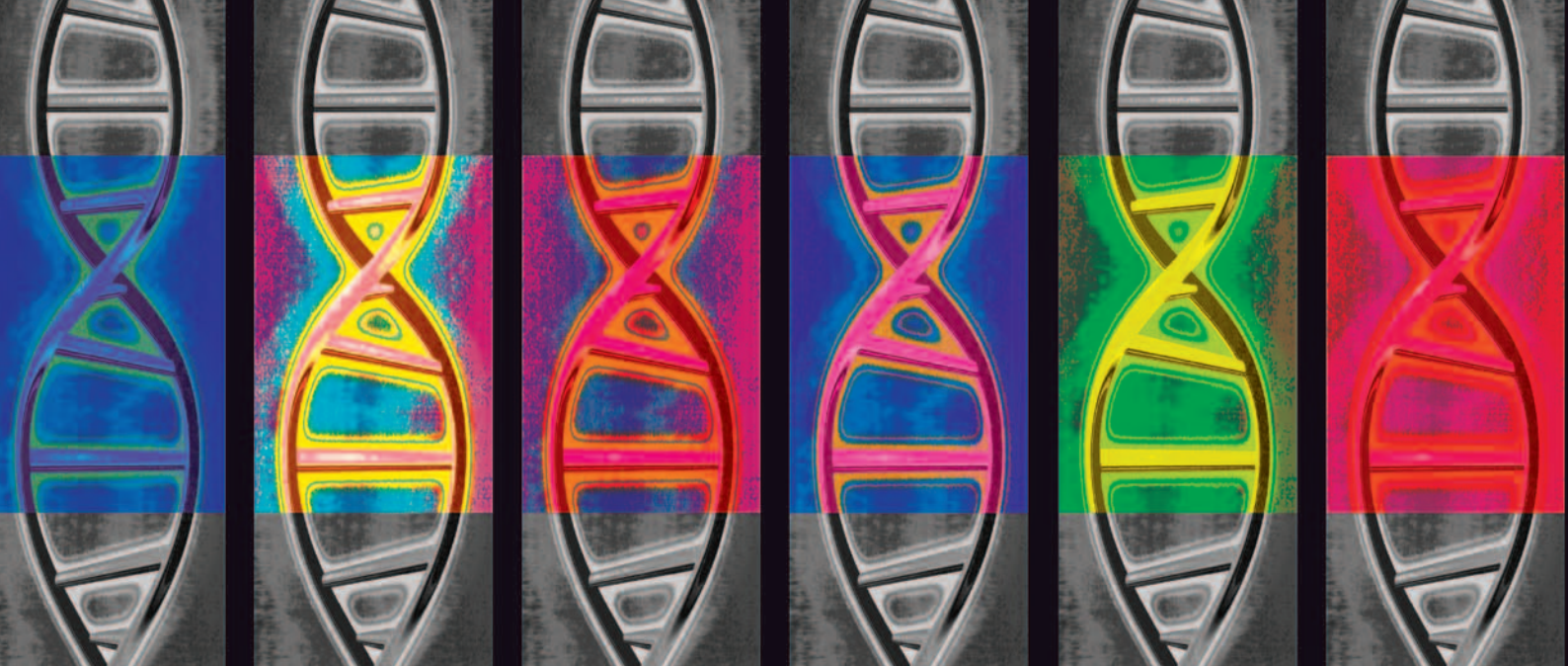
environment, which continually changes, is reflected in epigenetic changes, inflammatory cells and the paracrine mediators they produce, as well as oxidative stress. All of these nongenetic parameters in turn can have profound effects on the structure and function of the genome. The ultimate justification for a preventive approach to control of cancer is that cancer prevention is an opportunity to provide a higher quality of symptom-free and pain-free life to people, rather than waiting until someone has invasive and metastatic cancer with all of its attendant suffering for both patient and family.

We can look to the past for guidance. Fifty years ago, a unique spirit of intense cooperation found cures for acute childhood leukaemia and Hodgkin's disease, two previously fatal conditions. This triumphant work was achieved through highly collaborative efforts that tested multiple combinations of drugs. In the case of leukaemia, it took many years of research at multiple institutions to find the proper mix and dosage of vincristine, amethopterin, 6-mercaptopurine and prednisone (VAMP) that eventually enabled medical researchers Tom Frei and Emil Freireich and their team at the National Institutes of Health (NIH) to find a truly effective combination therapy. This effort involved not only the NIH, but teams at universities, as well as at several major pharmaceutical companies. A similar multi-group effort shortly thereafter led to the conquest of Hodgkin's disease.

Although we still have many cooperative groups for clinical trials, the all-hands-on-deck spirit that promoted the cure of childhood leukemia and Hodgkin's disease has largely disappeared in an increasingly competitive world. Regulatory and legal issues, as well as academic competitiveness and companies' perceived need to protect intellectual property, impede cooperation. To make substantial progress toward cancer prevention, we need to regain this lost ethos. ■

**Michael B. Sporn** is professor of pharmacology at Dartmouth Medical School in Hanover, New Hampshire.  
e-mail: michael.b.sporn@dartmouth.edu

1. Jemal, A. *et al.* *Cancer J. Clin.* **60**, 277–300 (2010).
2. Pleasance, E. *et al.* *Nature* **463**, 184–190 (2010).
3. LaCroix, A. *et al.*, *J. Natl. Cancer Inst.* **102**, 1706–1715 (2010).
4. Veronesi, U. *et al.*, *Ann. Oncol.* **17**, 1065–1071 (2006).
5. Andriole, G. *et al.* *New Eng. J. Med.* **362**, 1192–1202 (2010).
6. Schmidt, C. J. *Natl. Cancer Inst.* **102**, 1690–1692 (2010).
7. Meyskens, F. *et al.* *Cancer Prev. Res.* **1**, 32–38 (2008).
8. Albini, A. & Sporn, M., *Nat. Rev. Cancer* **7**, 139–147 (2007).
9. Sporn, M. *Nat. Clin. Pract. Oncol.* **3**, 364–373 (2006).
10. Issa, J. & Garber, J. *Cancer Prev. Res.* **4**, 6–8 (2011).



## EPIGENETICS

# Unravelling the cancer code

*There's more to the genetic causes of cancer than sequence mutations. This added complexity could offer scientists an opportunity to tackle cancer even earlier.*

BY VICKI BROWER

What causes a cell to turn cancerous? Ever since the discovery of oncogenes in 1989, the prevailing theory is that mutated genes drive the process: a predisposition to cancer is somehow written in our genetic code. This depressing portrait might have made cancer prediction easier, but hindered attempts at cancer prevention. Fortunately, the story became a lot more complicated.

It was geneticist Bert Vogelstein at Johns Hopkins University in Baltimore, Maryland, who clarified the role of that first oncogene as well as several others that followed; he is one of the originators of the notion of cancer as a genetic disease. Recently, while examining biopsies from ovarian cancer patients, he discovered that more than half of the tumour samples had mutations in the *ARID1A* gene<sup>1</sup>. Yet *ARID1A* does not directly stimulate the formation of tumours. "What we did not expect," says Vogelstein, "was that this gene is involved in determining the epigenetic changes that can lead to tumours." That is, the carcinogenic effect of *ARID1A* occurs by encouraging changes in gene expression levels and not in DNA sequence.

Vogelstein's findings were the tip of the iceberg. The Cancer Genome Project, an international effort to sequence the genomes of a number of different cancer types, has also found that "an incredibly high number of mutations actually affect epigenetics and

epigenetic regulators", says Jean-Pierre Issa, an epigeneticist at the MD Anderson Cancer Center in Houston, and a researcher on the project. And Vogelstein's insight into *ARID1A* is a clue to the pathways that connect genetics and epigenetics in cancer. "This closes the loop, as we find that even genetic lesions are controlled by epigenetic mechanisms," explains Issa.

Such findings are leading scientists to realize that the root cause of cancer is more complicated than inherited or acquired genetic mutations. "We used to think of cancer in binary terms — genetic mutations, or no mutations — but that's not the case anymore," says David Sidransky, an oncologist at Johns Hopkins. Researchers are finding that epigenetic changes frequently precede and can induce genetic mutations that cause cancer. If these early epigenetic alterations can be detected and reversed, it might be possible to prevent certain cancers.

## BEYOND THE HUMAN GENOME

The best known epigenetic modification is methylation — whereby a methyl group (CH<sub>3</sub>) attaches to a portion of DNA. Methylation of a gene reduces or stops its expression (see DNA methylation patterns, page S13). Patterns of methylation can be inherited from the mother or acquired during life. If the genetic code is the hardware for life, the epigenetic code is software that determines how the hardware behaves — and as such it can be rewritten.

The Human Genome Project (HGP) aimed

to decode life's hardware. Scientists hoped that one result would be discoveries of disease-causing genetic mutations. But mutations like those to *BRCA1/2*, and are highly predictive of cancer risk, were found to be the exception rather than the rule. Indeed, the HGP helped confirm that underlying most common diseases are hundreds if not thousands of genetic mutations that vary from person to person. "Even before the rough draft of the HGP was completed in 2000, it had become clear that it is impossible to understand the genetics of cancer without epigenetics," says Issa.

The HGP drew attention away from the nascent study of cancer epigenetics, which began in the mid-1980s at Johns Hopkins in the lab of cancer biologist Stephen Baylin. He noticed that cancer cells contained regions of DNA with increased methylation and hypothesized that if a tumour suppressor gene was hypermethylated, its activity would decrease or stop entirely — just as if it were a genetic mutation — allowing the tumour to flourish. In other words, Baylin reasoned, this epigenetic change would produce the same result as a genetic mutation.

Firm evidence came in 1994. Baylin and his colleague, oncologist James Herman, were investigating renal cell carcinoma (RCC),

the most common type of kidney cancer in adults. Around 60% of RCCs are caused by an inherited mutation in the von-Hippel Lindau

**NATURE.COM**  
Cancer epigenetics reaches mainstream oncology.  
[go.nature.com/5vi6bj](http://go.nature.com/5vi6bj)

MATT RAY/ENVIRONMENTAL HEALTH PERSPECTIVES



tumour-suppressor gene (VHL), which hobbles the gene's ability to express the tumour suppressing protein. Baylin and Herman showed that 20% of the remaining patients with the non-inherited form of RCC did not have a mutation in VHL. Their genes were silenced not by a mutation but rather by hypermethylation<sup>2</sup>.

The following year, in collaboration with Sidransky's lab at Johns Hopkins, Baylin and Herman showed that human cancers commonly arise when a particular tumour suppressor gene, known as p16, is hypermethylated. Moreover, in many cancers including RCC, epigenetic and genetic mutations often work in tandem: one of the two copies of a tumour suppressor gene is inactivated by genetic mutation, while the other is hypermethylated. This finding "convinced us that epigenetic abnormalities could play an important driving role in cancer — and we and many others have been pursuing this possibility ever since," says Baylin.

The move from a purely genetic to an epigenetic model is crucial for prevention strategies. As numerous gene therapy trials have shown, it is very difficult to treat a genetic disease by re-activating the dormant, mutated gene or by replacing it with a non-mutated one. "Epigenetic changes are reversible, and therefore have an edge over genetics," says Mukesh Verma, an epigeneticist at the National Cancer Institute's division of cancer control and population sciences in Bethesda, Maryland. Furthermore, epigenetic changes in cancer occur

before genetic mutations. "If you can prevent methylation of those tumour suppressor genes, you might have a valuable prevention strategy," says Baylin.

### THE ENVIRONMENTAL LINK

Epigenetics has also provided clues that link environmental factors with cancerous genetic changes. Changes in methylation can be detected in the blood of cancer-free individuals who smoke and eat high-fat diets, and these changes have been shown to precede genetic mutations<sup>3</sup>. More recently, Karl Kelsey, a molecular epidemiologist at Brown University in Providence, Rhode Island, has uncovered independent associations between epigenetic patterns in breast cancer tumours, the tumour size, alcohol consumption and folate intake<sup>4</sup>.

A prime candidate at the interface of environment and genetics is chronic inflammation, which is known to precede the development of numerous types of precancerous lesions — and indeed certain cancers themselves, including oesophageal, liver and colon cancers. Inflammation has been linked with increased DNA methylation in otherwise healthy looking tissue. Issa calls chronic inflammation "a truly epigenetic phenomenon".

Long-term inflammation may result from infection with *Helicobacter pylori* or hepatitis C virus, or from autoimmune diseases such as ulcerative colitis (a form of inflammatory bowel disease). People with ulcerative colitis often develop colon cancer at a younger age — for example in their 50s — than the 60 to 70 year average age of onset. "All of the epigenetic changes that occur in colon cancer, specifically DNA hypermethylation and gene silencing, are accelerated — and appear in the inflamed tissue prior to actual cancer," says Baylin.

Half the world's population is infected with inflammation-causing *H. pylori*, yet stomach cancer afflicts barely 0.03% of those. "There must be something in the body itself that makes it react differently to infection," says Emad El-Omar, a gastroenterologist at the University of Aberdeen in Scotland. El-Omar is investigating whether genetic variation can influence this response.

Genetic polymorphisms are normal genetic variations within a population that can subtly raise or lower each person's levels of a particular protein. While everyone within the normal population produces the same pro- and anti-inflammatory chemicals (or cytokines) an individual's particular levels vary according to genetic make-up. Certain polymorphisms, El-Omar hypothesized, might tip the balance towards chronic inflammation and then to cancer.

El-Omar found that polymorphisms in the inflammation-related IL-1B and transforming growth factor (TNF)- $\alpha$  genes determine the levels of circulating IL-1B and TNF- $\alpha$  pro-inflammatory cytokines. People with a genotype disposed to higher levels of these

cytokines have an increased risk of developing stomach cancer following *H. pylori* infection. El-Omar has made similar discoveries in other cancers. In colorectal cancer, for example, he discovered there is an inflammatory environment in and around premalignant lesions. Within this region he found nine differentially expressed genes linked with inflammation, including those responsible for IL-8 and the CXCL- family of cytokines<sup>5</sup>.

Others researchers are making similar connections. For example, there is evidence that pro-inflammatory prostaglandin endoperoxide synthase<sup>6</sup> and TGF- $\beta$ <sup>7</sup> are both significantly associated with increased risk of colon cancer. Although the evidence linking inflammation to epigenetics and cancer is mounting, the underlying mechanism of the association and the value of screening for these polymorphisms remain less clear.

### POTENTIAL FOR REVERSAL

Drugs and dietary substances that alter epigenetic pathways are currently being tested. During his research on RCC, for example, Baylin and colleagues were able to reverse hypermethylation of the VHL gene with the drug 5-azacytidine. Trials of demethylating drugs as adjuvant treatments to prevent lung cancer recurrence are underway. If successful, prevention trials are the next logical step. "We need five- and ten-year survival data with current drugs to be sure there are no secondary effects before we give them to reasonably healthy people for prevention," says Issa. He sees a different source for the first wave of preventive medications. "I would bank on discovering more 'gentle' approaches to epigenetic manipulation for cancer prevention — be they natural products, existing drugs with a good safety records, or even vitamins or diet."

Epigenetic changes also have potential utility as biomarkers (see Portents of malignancy, page S19). Being able to read the methylation pattern of certain genes could give scientists a heads-up for people on the brink of developing cancer.

Slowly the importance of the epigenome in cancer development is being appreciated. "Geneticists are hugely more aware of the importance of epigenetics in the development of cancer," observes Baylin. When it comes to cancer prevention, the future could lie in arresting the reversible epigenetic changes before irreversible mutations take hold. ■

**Vicki Brower** is a freelance writer in New York.

1. Jones, S. *et al.* *Science* **330**, 228–231 (2010).
2. Herman, J.G. *et al.* *PNAS USA* **91**, 9700–9704, (1994).
3. Brait, M. *et al.* *Cancer Epidemiol Biomarkers Prev.* **18**, 2984–2991 (2009).
4. Christensen, B. C. *et al.* *PLoS Genetics* **6**(7):e1001043 (2010).
5. McLean, M. H. *et al.* *PLoS One* **6**(1), e15366 (2011).
6. Frank, B. *et al.* *Cancer Genetics* **127**, 2822–2830 (2010).
7. Slattery, M. *et al.* *Cancer Epidemiol. Biomarkers Prevention.* **20**, 57 (2011).

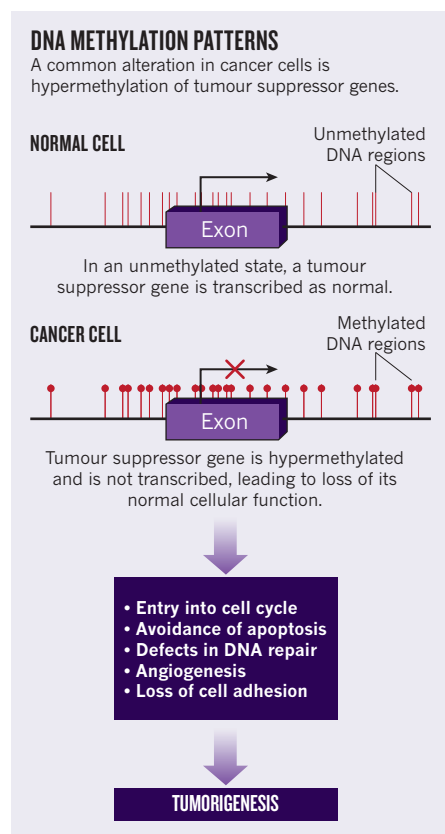






ILLUSTRATION BY ALBERTO RUGGERI

supplant today's colonoscopies, mammograms and pap smears. If some of these approaches can be shown to prevent cancer deaths and cut costs, they stand a good chance of becoming part of regular patient care.

### LIGHT PROBES

Many researchers are trying to improve on existing techniques such as endoscopy, delivering images from inside the body through fiber optics. Engineers at Duke University, North Carolina, for instance, have designed an optical system to search for premalignant changes in patients with Barrett's esophagus, in which stomach acid alters the cells lining the esophagus. The condition more than doubles the risk of esophageal cancer. Unlike conventional endoscopy, the Duke technique, called angle-resolved low-coherence interferometry, images structures beneath the surface of a cell for a sort of optical biopsy. Adam Wax, one of the Duke engineers, says looking at the basal layer of the epithelium, about 300 micrometers beneath the surface, seems most diagnostically useful. The system splits infrared light into two beams, and compares how far each travels to determine how deep it penetrates into the cell. Measuring the angle at which light bounces off cellular structures reveals the size of structures at increasing depth. The resolution is high enough to distinguish a normal-sized nucleus, about 10 micrometers in diameter, and a larger, precancerous one at least 13 micrometers.

Wax says his enhanced endoscopy could provide better targets for biopsies — and, eventually replace biopsies altogether. According to the NCI, esophageal cancer causes nearly 15,000 deaths in the United States each year. "We hope that by contributing this tool we'll be able to shift that number downwards — the way it's gone with colonoscopy," says Wax, who has launched a company, Oncoscope, to raise funds for clinical trials.

A similar light-based technique, optical coherence tomography (OCT) — could detect non-melanoma skin cancer below the surface of the skin, where standard visual exams can't see. Where Wax aims to get a precise measurement of cell structures, OCT provides images that doctors can examine. OCT — already used by ophthalmologists to examine the inside of the eye, also uses interferometry to image intracellular structures so doctors can see if they're abnormal. A British company, Michelson Diagnostics, is developing a hand-held OCT scanner to detect non-melanoma skin cancer below the surface of the skin.

"We're very good at seeing where the lesion is," says biomedical engineer Gordon McKenzie, Michelson's medical applications director. "What we're doing now is gathering the evidence of whether we're seeing a cancer or a precancer." He says the machine, VivoSight, is comparable in both appearance and cost to the ultrasound machines found in obstetricians' offices. He hopes that the device, now

### EARLY DETECTION

# Spotting the first signs

*The sooner a cancer is found, the better. New technologies are pushing the limits of detection — towards the frontier of prevention.*

BY NEIL SAVAGE

One day, a few years hence, a patient having a routine check-up might do little more than blow into a small machine at the doctor's office and, within a couple of minutes, be told whether there are any early signs of cancer. For another patient, a routine blood test to monitor cholesterol might present an opportunity to check for stray cells from tumours too small to spot. A dermatologist, instead of eying a mole and perhaps slicing it off to biopsy, could instead peer at it through a machine to instantly tell whether it is malignant or benign.

These, at least, are the visions of researchers

developing technologies to detect the early signs of cancer. Better screening — looking for signs of cancer in people with no symptoms, as opposed to diagnosing suspected cancer — increases the odds that doctors will find cancer at its earliest stages when chances of a cure are higher. Screening has already reduced cancer deaths: the US National Cancer Institute (NCI) estimates that colonoscopies can lower mortality from colorectal cancer by at least 60%, and the National Lung Screening Trial recently found that computed tomography scans of heavy smokers could cut lung cancer deaths by as much as 20%. Researchers are exploring a new suite of potential screening methods that could one day join or even

in clinical trials, will reduce the number of biopsies of non-malignant lesions, and make the biopsies that are performed more accurate. Right now, he says, as many as 80% of patients who have a malignant mole removed need more tissue excised after tests find that some cancerous tissue may have been left behind. Eventually, he says, their scanner might help find new lesions in the colon and cervix as well as the skin.

### NANOPARTICLES AND MOLECULES

Telling the precancerous from the cancerous is useful, but the earlier it is possible to detect aberrant cellular changes, the closer that comes to being equivalent to prevention. "Diagnosis is very nice, but screening is much more important," says Hossam Haick, a chemical engineer at Technion Israel Institute of Technology in Haifa. Haick has built an array of nanosensors that, he says, detect biomarkers of certain cancers years before they develop. He bases this assertion on histological studies showing that cells producing these molecules eventually become cancerous.

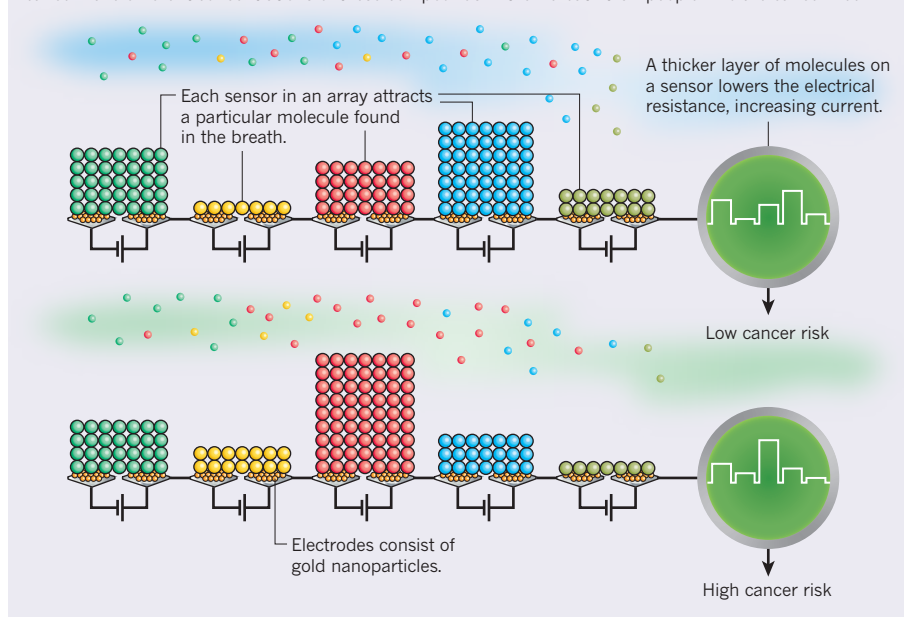
Haick's nanosensors measure volatile organic compounds released by cells into the bloodstream and then exhaled — a sort of cancer breathalyzer. The array consists of about 40 gold nanoparticles attached to molecules sensitive to various organic compounds such as alcohols, benzenes and alkenes. The patient blows into a bag, and as the breath passes through the array any volatile organic compounds bind to their complementary sensors and emits a signal. Software probes the pattern of any signals to get a signature of a type and stage of cancer (see Telltale cancer breath). By testing patients with known cancers, Haick identified signatures for lung, breast, colorectal and prostate cancer. He is working on finding signatures for ovarian, liver and gastric cancer, and has formed a company to commercialize the technology.

One challenge in early detection is picking up a hint of cancer from tumours still too small to show up under current screening techniques. Mehmet Toner, a biomedical engineer at Harvard Medical School and Massachusetts General Hospital, is building a sensor chip that can find a single circulating tumour cell (CTC) in a billion or more blood cells. A blood sample is injected into the chip, along with antibodies that bind to proteins on the surface of the cancer cells. Also attached to the antibodies are tiny magnetic beads. Once suspect cells are tagged with those beads, they can easily be collected by other magnets for closer examination. Toner says the chip, which he is developing with Johnson and Johnson subsidiary Veridex, should prove to be far more sensitive than previous techniques.

In addition to early detection, Toner says, the blood test is attractive because it's easier — and cheaper — than something like a colonoscopy. Colonoscopy is recommended for

### TELLTALE CANCER BREATH

Everyone's breath contains certain chemicals (e.g., alcohols, benzenes and alkenes). People with early stage cancer have different concentrations of these compounds in their breath than people who are cancer free.



everyone over 50 years of age, but Toner says the CTC chip could act as a screen to rule out some patients — and perhaps to rule in younger people if the chip raises any issues. The aim of the Veridex collaboration is to manufacture 10,000 to 20,000 of these magnetic-bead chips over the next two years, then use them in clinical trials to see how well they detect CTCs. Toner notes that it is still an open question whether these cells, undetectable by current methods, will ever develop into cancers that can kill. Whereas today's techniques may find cancer too late, an overly sensitive test might cause unnecessary alarm. "It could be that cancer is something we all live with, but [current] tests are such that we only look at it when it becomes a problem," he says.

### SEARCH FOR RELEVANCE

In fact, such a risk of overdiagnosis is a major potential problem with early detection technologies, says Sanjiv Sam Gambhir, at the Canary Center at Stanford for Cancer Early Detection. "Our goal isn't to find all cancers early," he emphasizes. "Our goal is to find early relevant cancers — that is, cancers that will go on to kill." That will require biologists to work along with the technology developers — both to understand the significance of what the tests find and to identify new biological targets to refine the detectors. Gambhir, for instance, talks about identifying prognostic biomarkers — proteins that appear only in patients with the most aggressive tumours.

Gambhir specializes in molecular imaging, in which a molecule that binds to a specific biological target is labelled, often with a radioisotope, so that it shows up more clearly in existing imaging technologies. His group

created a probe that binds to tumour cells, but instead of a radioactive label, they attached a microbubble — a tiny, gas-filled sac that reflects sound waves and shows up brightly in an ultrasound image. Another approach relies on Raman light scattering, in which a small fraction of the photons reflecting off a surface shift their wavelength in a specific way. Gambhir injects probes labelled with gold nanoparticles, then uses a colonoscope to find their Raman signal. This technique could improve the sensitivity of colonoscopy, which is good at identifying polyps but tends to miss flat precancerous lesions.

Although the ultimate goal is to develop techniques to screen the general population for early signs of malignancy, it is easier to tell if the new tests work when researchers already know a patient has cancer. Thus many techniques are first being developed for use in the easy cases — staging tumours or identifying sites for biopsy in patients with known malignancies. The technological progression, says Toner, is to go "from metastatic cancer to diagnosed but early-stage cancer to high-risk groups free of cancer to the wider population". Once researchers prove that the new technologies work in that wider population, physicians will have at their disposal a powerful set of early warning systems that can give us bad news in time for it to be not the worst news. ■

Neil Savage is a writer based in Massachusetts.

1. National Lung Screening Trial Research Team, *Radiology* **258**, 243–253 (2011).
2. Wax, A. et al. *Gastroenterology* **140**, 42–50, (2011).
3. Peng, G. et al. *British Journal of Cancer* **103**, 542–551 (2010).
4. Willman, J.K. et al. *J. Nuclear Medicine* **51**, 433–440 (2010).





ILLUSTRATION BY MATTHEW HOLLISTER

men). In 2003, the ACS estimated that excess weight could account for one in seven cancer deaths among men and one in five among women in the United States.

Since 2003, the evidence linking body weight and cancer has grown stronger. One of the most comprehensive efforts to examine this relationship came from the World Cancer Research Foundation and the American Institute for Cancer Research in Washington DC. At UCL, Marmot led a team of 21 scientists to review some 7,000 related studies, and again found a convincing link between excess weight and many cancers. They found a link between excess body fat and a variety of cancers. They recommended that people strive to be as lean and active as possible, and that they avoid sugary drinks.

Researchers have yet to fully understand how being overweight can cause cancer. The mechanism likely depends on the type of malignancy. For example, abdominal fat presses on the stomach, causing acid to splash up into the oesophagus. That acid leads to tissue damage, which can lead to oesophageal cancer. Oestrogen, produced by fat cells, appears to play a role in endometrial cancer and breast cancer in postmenopausal women. "Obese women have about three times the circulating level of oestrogens as lean women," says Walter Willet, a prominent nutrition researcher at Harvard University.

Obesity can also cause the body to become less responsive to insulin. To compensate, the pancreas churns out more of this potent growth factor. Researchers posit that excess insulin can cause cancer cells to proliferate. Diabetics treated with the drug metformin, which lowers insulin levels, appear to have a lower risk of many cancers, including pancreatic and breast. It is unclear, however, whether metformin's anti-cancer activity is related to insulin<sup>2</sup>.

Recent research has confirmed the complementary effect: losing weight can make a person less prone to cancer. Researchers in Sweden, for example, tracked two groups of 2,000 overweight men and women and found that bariatric surgery cuts the risk of cancer in women by 42% (ref. 3). Another study found that gastric bypass cut the risk of developing cancer by 24% and the risk of dying of cancer by 46% (ref. 4). (In both studies, this trend held only for women, not men, which suggests weight loss may have a particularly large impact on breast and endometrial cancers<sup>5</sup>.)

Regular exercise can cut cancer risk as well—and not just because it often leads to reduced weight<sup>6</sup>. "We've been able to tease out the individual effects of being physically active versus being overweight," says Christine Friedenreich, a cancer epidemiologist and leader of population health research at University of Calgary in Canada. One theory is that active people tend to digest food faster. "Being physically active may decrease

## LIFESTYLE

# Breaking the cancer habit

*It's the simple things in life that sometimes mean the most to people—and do the most good.*

BY CASSANDRA WILLYARD

**D**rug companies spend billions of dollars developing high-tech therapies to deploy against cancer. Yet millions of people still die of the disease each year. Preventing cancer appears to be a far simpler proposition. The most effective steps to curb cancer are low-tech: get people to stop smoking and lose weight. Smoking, long known to be a risk factor, and obesity, more recently recognized as one, together account for roughly half of all cancer cases. But as anyone who has tried to kick the habit or lose a few pounds knows, both steps are easier said than done.

To have a real impact on cancer, health officials need to address the "causes of causes of cancer," says Michael Marmot, an epidemiologist at University College London (UCL). That means finding new ways to curb smoking and obesity. "Simply telling people not to get fat is not very effective," says Marmot. Policies that

make unhealthy lifestyles more inconvenient and expensive, while making healthier ones easier and cheaper, can have a large impact. So can medical interventions. For example, researchers are developing new ways, such as the nicotine vaccine, that might help smokers quit for good.

## EAT LESS, MOVE MORE

Many studies in the past couple of decades have examined the impact of dietary factors on cancer risk (see The omnivore's labyrinth, page S22). How much we eat may be as important as what we eat. But overeating may be more important than choosing the right foods. In a study that followed 900,000 adults for 16 years, researchers at the American Cancer Society (ACS) found a significant association between body mass index and higher mortality owing to cancers of the oesophagus, colon, liver, gallbladder, pancreas, kidney, breast (in women), uterus, cervix, ovary, prostate and stomach (in



the chance of any carcinogenic products that happen to be going through the colon to have contact with the mucosal lining," she explains. Similarly, better lung function in fit individuals could limit their exposure to airborne carcinogens. Hormones might also be a factor. Friedenreich and her colleagues found that postmenopausal Canadian women who walked for about three hours a week had lower oestrogen levels after a year than sedentary women<sup>7</sup>.

## HEALTHY INFLUENCES

Providing consumers with better nutritional information may help them choose a more nutritious diet and avoid the obesity that raises their cancer risk. Toward that end, David Katz, director of the Yale-Griffin Prevention Research Center, has developed an algorithm for ranking foods according to their overall healthiness. The formula, NuVal, accounts for ingredients including salt, vitamins, saturated fat, fibre, sugar, cholesterol, protein, as well as overall calorific value. It assigns each food a score between one and a hundred<sup>8</sup>. The higher the value, the healthier the food (pineapple 99, butter cookie 1). According to Katz, NuVal has been used to score more than 90,000 foods.

Economic incentives can also change nutritional behaviour. In January 2011, Wal-Mart, the largest grocer in the United States, pledged to lower the price of fruits and vegetables, cut salt and sugar content in packaged foods, and stop selling any foods containing trans fats within the next five years.

In 2010, the US Department of Agriculture launched a pilot study to see if cutting the price of fruit and vegetables would prompt low-income families to eat more of them and reduce their consumption of less healthy foods. The 1,500 participating families will get 30 cents added to their food benefit card balance for every dollar they spend on fruits and vegetables. The study is set to begin in late 2011 and wrap-up in 2013.

An alternative strategy would be to tax unhealthy foods, like soda and sugary sports drinks. Denmark already has such a tax. "They put a very high tax on sugar-sweetened beverages, a medium tax on diet beverages, and no tax on water and low-fat milk," says Barry Popkin, a nutritionist at the University of North Carolina's Interdisciplinary Center for Obesity.

A study published in February 2011 hints that taxing the unhealthy foods might be the best approach. Researchers at the State University of New York at Buffalo recruited 42 mothers to shop at a simulated supermarket stocked with pictures of everything from whole wheat bread to bananas to carbonated sweet drinks. Each participant was given \$22.50 and asked to select a week's worth of food. During five 'shopping trips', the researchers manipulated the prices of the foods, first charging prices comparable to a local grocery store, then lowering the price of healthy foods by 12.5% to 25%, and then hiking the prices of unhealthy foods by roughly the same amount. Result: raising the price of junk food lowered the total calories purchased. Healthy food subsidies, on the other hand, increased the total amount of fat, protein and carbohydrates bought. The participants spent the money saved to buy junk food<sup>9</sup>.

Neither taxes nor subsidies will end the obesity epidemic. But just because an intervention doesn't lead to weight loss doesn't mean it's a dud, says Katz. He compares obesity to a flood. Each intervention, he says, is a sandbag in a much-needed levee. "No one of them by itself can stop the flood," he says. "Only when we've done enough of these things in enough places will they add up to be a levee that's higher than the floodwaters."

Smoking still accounts for a third of all cancer deaths. And tobacco, unlike most foods, is addictive. So kicking the habit often requires medical intervention.

Most over-the-counter smoking-cessation

therapies, such as the nicotine patch and nicotine gum, curb the symptoms of withdrawal by providing small doses of nicotine. Alternative strategies aim to make smoking less addictive. Varenicline, for example, a drug approved in 2006 and marketed as Chantix, partially binds to the nicotine receptor in the brain. It is designed to block nicotine, and also to partially activate the receptor. The idea is to prevent smokers from getting a rush if they smoke, but to give them enough dopamine to help curb cravings. And NicVax, an anti-smoking vaccine in phase III trials, prompts the body to raise an immune response against nicotine. "When someone smokes, the antibody attaches itself to the nicotine molecule," says Dorothy Hatsukami, a specialist in tobacco addiction at the University of Minnesota's Masonic Cancer Center. Bound together, the two molecules are too large to penetrate the blood brain barrier. "It reduces the level of nicotine that can enter the brain at any one time," says Hatsukami, who is leading one of the NicVax trials.

Anti-smoking regulations could have an even greater impact on cancer. In 2009, US lawmakers gave the Food and Drug Administration (FDA) unprecedented power to regulate tobacco. The FDA now prohibits all flavourings, with the exception of menthol, and requires tobacco product manufacturers to register their products with the FDA. The statute also gives the FDA the power to limit the amount of nicotine in tobacco products. In theory, the agency could cut nicotine to levels that would "render the products less or non-addictive," says Clifford Douglas, a tobacco control specialist who heads the University of Michigan Tobacco Research Network. Such a restriction could have an enormous impact. "If one wants to cut seriously into the tobacco epidemic, they must deal with nicotine." Local regulations can play a role as well. Cities all over the world have banned smoking in bars and restaurants, but some are going a step further. In February 2011, city councillors in New York City voted to ban smoking in public parks, beaches and boardwalks.

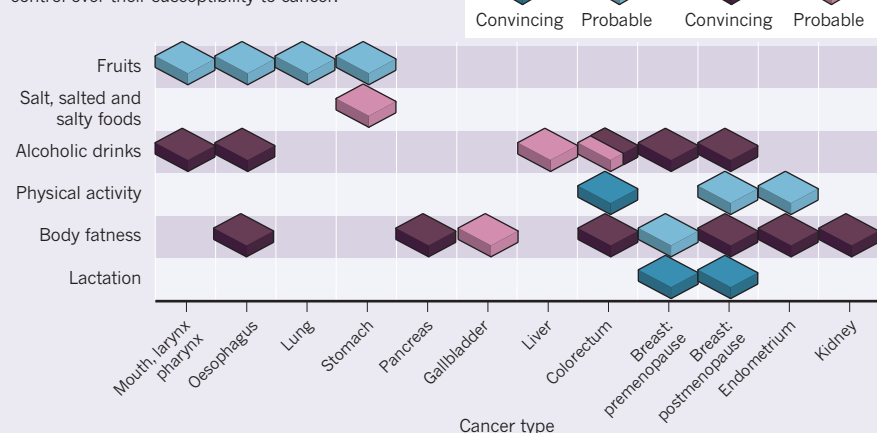
With respect to both obesity and smoking, part of the battle involves convincing people that much cancer is, in fact, avoidable. That could prove challenging, Marmot says. "I think most people think cancer is an act of God or Darwin."

**Cassandra Willyard** is a freelance writer in New York.

1. Calle, E. E. *et al.* *N. Engl. J. Med.* **348**, 1625–1638 (2003).
2. Pollak, M. *Cancer Prev. Res.* **3**, 1060–1065 (2010).
3. Sjöström, L. *et al.* *Lancet Oncol.* **10**, 653–662 (2009).
4. Adams, T. D. *et al.* *Obesity* **17**, 796–802 (2009).
5. Ashrafian, H. *et al.* *Cancer Nov 29*. [Epub ahead of print] (2010).
6. Friedenreich, C. M. *et al.* *Eur. J. Cancer* **46**, 2593–2604 (2010).
7. Friedenreich, C. M. *et al.* *J. Clin. Oncol.* **28**, 1458–1466 (2010).
8. Katz, D. L. *et al.* *Am. J. Clin. Nutr.* **91**, 1102S–1108S (2010).
9. Epstein, L. H. *et al.* *Psychological Science* **21**, 406–414 (2010).

## RISK AND REWARD

Through diet and exercise, individuals can take some control over their susceptibility to cancer.





B.S.I.P. / BOUCHARLAT / SCIENCE PHOTO LIBRARY

PERSPECTIVE

# Tackling the real issues

Successful prevention requires attacking the causes, says **Stephen S. Hecht**—and the main target remains tobacco.

The best way to control cancer is to prevent it from happening. Data from the United States indicate that the age-standardized incidence of cancer continues to decrease in this century, mainly owing to decreases in the major cancers: lung, prostate and colorectal in men; breast and colorectal in women<sup>1</sup>. Similar trends exist in other Western countries, although not in less developed or transitioning ones<sup>2,3</sup>. These decreases are largely attributable to cancer prevention. Millions of lives have almost certainly been saved.

Unlike prevention of heart disease, which is now a commonly understood goal of a healthy lifestyle, the concept of cancer prevention gets little attention. Curing cancer is newsworthy and glamorous. Prevention is not. One can meet the survivors of childhood leukaemia or breast cancer and marvel at their good fortune. It is difficult to write an engrossing story of cancer about a person who did not get cancer in the first place.

We now know a great deal about the causes of common cancers — and the better we understand cause, the more able we are to conceive of effective preventive measures. Geographic and economic differences in cancer incidence and mortality are striking. Common forms of cancer vary greatly between high-income countries and low- to moderate-income countries<sup>2,3</sup>. Wealthy countries are beset by high incidences of lung, breast, prostate and colorectal cancers. Migrants from one geographical area to another adopt the risk factors of their new

homes. Furthermore, the lifestyle habits of certain groups, such as Mormons and Seventh Day Adventists, lead to significantly lower cancer rates, in part because of their abstinence from tobacco and alcohol<sup>3</sup>. These facts unmask the role of lifestyle factors in the common cancers and illustrate the huge potential of cancer prevention (see Breaking the cancer habit, page S16).

Cigarette smoking, the cause of 90% of lung cancer, stands out as one of the best examples in which cause and prevention are intimately linked. Smokers are exposed to multiple DNA-damaging carcinogens and consequently have acquired multiple mutations in genes that control cellular growth<sup>4</sup>. They are also exposed to multiple tumour promoting substances and inflammatory agents that exacerbate the process. The addictive power of nicotine perpetuates the persistent exposure to these toxicants.

Effective tobacco control — led by clean air legislation, taxation, and anti-tobacco advertising — is contributing to decreased lung cancer incidence and in some Western countries. We can expand our preventive activities against tobacco-related cancer by achieving a better understanding of the biochemical, genetic and behavioural mechanisms of smoking. This knowledge could help us identify people who have a particularly high susceptibility to these cancers; we could then target those individuals for new prevention measures, such as a nicotine vaccination and chemoprevention.

We know less about major causes of breast, prostate and colorectal cancer. Estrogen is clearly involved in breast cancer aetiology and androgens are critical in prostate cancer development. Consumption of red meat cooked at high temperatures plays a role in colorectal cancer. We must expand our understanding of the causes of these cancers and translate it to cancer prevention strategies.

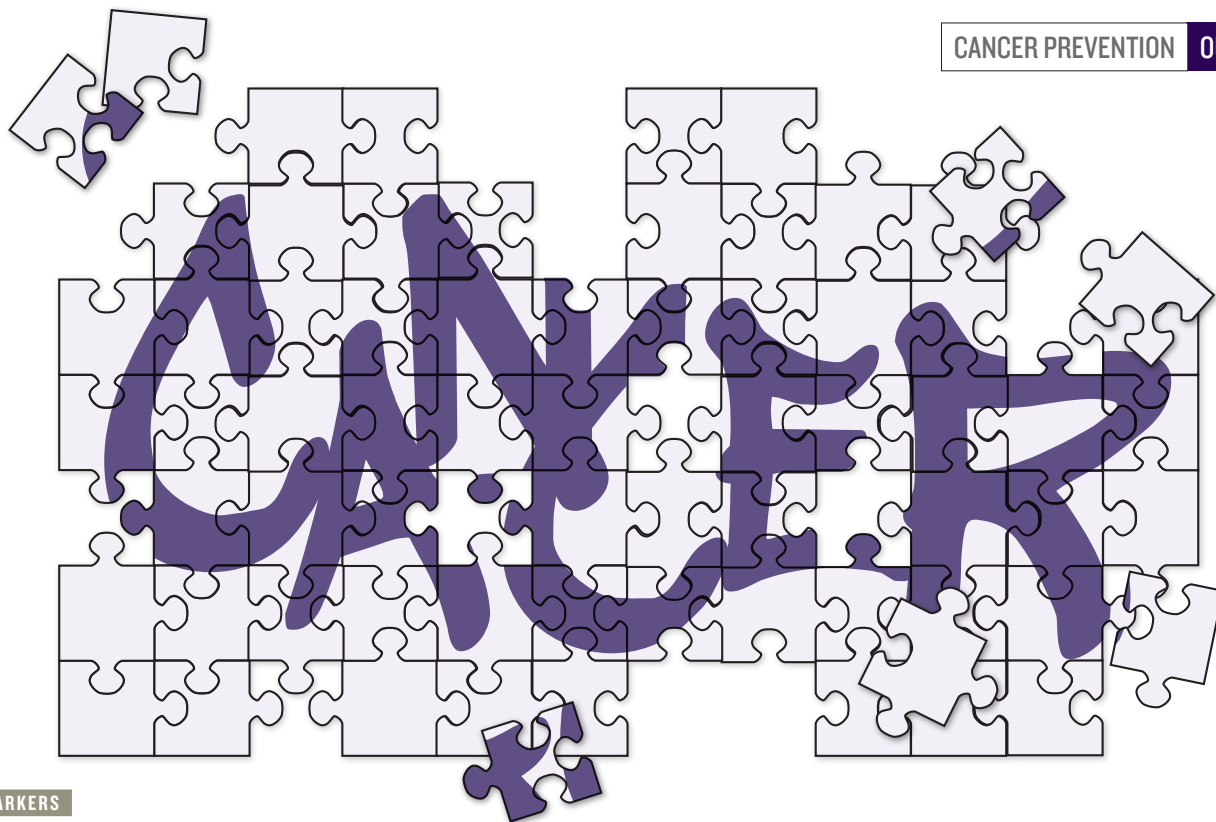
There are other significant and successful strategies for primary and secondary cancer prevention: vaccination and screening for cervical cancer; vaccination against hepatitis B and avoidance of aflatoxins for liver cancer; avoidance of excessive sunlight exposure for skin cancer; limiting alcohol consumption for head and neck cancer, and liver cancer.

Education and public outreach are critical. Here, we might emulate our colleagues in cardiovascular disease prevention research. There is a sharply increased awareness of the preventive power against heart disease of low-fat diets, cholesterol reduction (in part owing to statins), treatment of hypertension, exercise, and avoidance of tobacco. This awareness, along with improved medical care, has led to a significant reduction in deaths from heart disease — from about 320 deaths per 100,000 people (younger than 85 years) in 1975 to 130 deaths per 100,000 people in 2006. This almost 60% drop in mortality rate far exceeds the 11% decrease in the rate of cancer deaths (from 180 to 160 per 100,000 people) over the same period<sup>1</sup>.

As we allocate research resources on the life-saving potential of cancer prevention, we need to keep in mind the prime importance of lifestyle factors in cancer aetiology, and continue to support studies to better understand the specific causes and mechanisms of common cancers. We need to identify susceptible individuals and target them for preventive interventions, including chemoprevention using non-toxic or dietary agents with demonstrated efficacy. We should not be distracted by fleeting, flamboyant approaches that take us off the main task of dealing with the lifestyle factors that link cause and prevention. ■

**Stephen S. Hecht** is Wallin Professor of Cancer Prevention at the University of Minnesota's Masonic Cancer Center.  
e-mail: hecht002@umn.edu

1. Jemal, A. et al. *CA Cancer J Clin* **60**, 277–300 (2010).
2. Jemal, A. et al. *Cancer Epidemiol Biomarkers Prev.* **19**, 1893–1907 (2010).
3. Boyle, P. & Levin, P. World Cancer Report, 2008. (IARC, 2008).
4. US Department of Health and Human Services *How Tobacco Smoke Causes Disease*, 221–353 (DHHS, 2010).

**BIOMARKERS**

# Portents of malignancy

*Being able to determine an individual's chances of developing cancer will greatly improve risk management strategies and recruitment to clinical trials.*

BY VICKI BROWER

**B**efore deciding to light your first cigarette, imagine knowing your exact chance that this will lead to cancer. Although smoking is the main cause of lung cancer, only 10%–20% of smokers and former smokers actually develop the disease. The reasons for this — and for the changes that lead to many cancers — have eluded researchers for decades.

Lung cancer, as the biggest cause of cancer death worldwide, is a priority of prevention strategies. And to make the biggest impact in cancer prevention, it is vital to target those individuals with the highest risk of developing cancer. This is where biomarkers come in. Individuals identified as being high risk using biomarkers could receive counselling for lifestyle changes (see *Breaking the cancer habit*, page S16), or they might be eligible for chemoprevention (see *First line of defence*, page S5). Even before that point, using biomarkers to select people for cancer prevention studies would allow for more powerful trials (see *Designing smarter cancer prevention trials*, page S20). Indeed, there isn't a clear distinction between the types of biomarkers that help identify risk before cancer has developed and those that indicate that cancer is just starting to grow.

The process of carcinogenesis takes years, if not decades. The search is on to discover and validate the often-subtle, microscopic changes in the constituents of the blood, sputum, urine or tissue samples that herald cancer. Prior to activated oncogenic pathways, there is the possibility of identifying sluggish DNA repair mechanisms, changes in gene expression, or detecting the low-level immune response to the presence of a nascent tumour.

This research activity is not without its controversies. There have been numerous false starts, where promising biomarkers did not stand up to rigorous testing. To be useful, a biomarker needs to have sensitivity — that is, the likelihood that it detects disease — of at least 90%. The other key quality is specificity — the probability that a positive signal is a true sign of disease and not an error. That, too, should be 90% or more for a biomarker to be of clinical value. Although no lung cancer biomarkers yet meet that 90/90 standard, there are several promising candidates.

## CHANGING EXPRESSION

Avrum Spira, a pulmonologist at Boston University, has been using bronchoscopy to brush cells from the bronchial airways of healthy smokers and non-smokers, followed by gene expression profiling to compare thousands of genes. He has found that the cells

lining smokers' airways show signs of genomic changes related to inflammation and cell proliferation, even when they appear normal with standard bronchoscopy. By comparing these results to those of smokers with suspected lung cancer, Spira's group has identified an 80-gene signature that could identify patients with early-stage lung cancer with about 90% sensitivity. This signature, says Spira, "is the proverbial canary in the coalmine".

However, bronchoscopy is an invasive procedure, so Spira broadened his search to more accessible parts of the airway. He recently found abnormal gene expression in epithelial cells of the nose and mouth that resemble those in the bronchial airway<sup>1</sup>. Analyzing these cells using a simple swab could serve "as a mass screening tool in population-based studies," he predicts.

Digging into the pathways underlying these precancerous changes, Spira, together with genomicist Andrea Bild from the University of Utah, found phosphatidylinositol 3-kinase (PI3K) signalling pathway, already known to be involved in the development of cancer<sup>2</sup>. By comparing PI3K activity levels in cells from apparently healthy smokers to those of smokers with mild-to-moderate abnormalities, they found that this pathway "might be activated before tumorigenesis," says Spira, making PI3K a prime biomarker candidate.



## DESIGNING SMARTER CANCER PREVENTION TRIALS

In April 2010, an international team of researchers from academia and drug company GlaxoSmithKline reported that dutasteride, a drug already approved for the treatment of benign prostatic hyperplasia, reduced the chances that men considered at high risk for prostate cancer would develop the disease. The four-year trial included more than 8,100 men and met the gold standard for clinical trials: it was randomized, double-blind, and placebo-controlled; it studied parallel groups at multiple medical centres; and it assessed outcomes with biopsies at two years and four years. In the end, men who took dutasteride were 23% less likely to have a positive biopsy for cancer than those on the placebo. GSK submitted this data in its application to the US Food and Drug Administration to market the drug for prostate cancer prevention



— this January, the FDA said No.

Although it is not unusual for the FDA to reject a drug application supported by apparently positive data, this case illustrates the particular challenges surrounding clinical trials for cancer prevention. When the aim is to decrease the incidence of cancer in large

populations, studies on preventive agents require large patient cohorts — sometimes approaching 20,000 participants — and take years or even decades to perform. This combination makes them especially unwieldy compared to tests with therapeutic compounds, which can much more quickly be seen to work,

or not, by testing them exclusively in people who have the disease. In cancer prevention drug trials, the usual gold standard barely rates a bronze.

Since preventives are intended for apparently healthy patients, trials require a high confidence that the anticipated anticancer benefit will outweigh any harmful side effects. In the dutasteride trial, statistical analysis showed that the decrease in cancer was driven mainly by a reduction in less serious tumours that might not even require treatment. In addition, men who took the drug were slightly more likely than those on a placebo to develop more aggressive tumours. The FDA's expert advisory panel concluded that the prevention benefits failed to outweigh this risk.

Researchers say two things are needed to decrease the length and size of prevention studies. One is

JAH CHIKWENDU/THE WASHINGTON POST VIA GETTY IMAGES

The PI3K pathway might also be used for chemoprevention. Early trials have shown that the administration of a compound that decreases PI3K activity causes regression of abnormal lesions in the airways of high-risk smokers<sup>3</sup>.

## DNA DAMAGE

As part of daily living, DNA frequently sustains damage. If not repaired, this can lead to mutations that replicate, resulting in abnormal and then cancerous growths. Certain mechanisms usually prevent this from occurring. The enzyme 8-oxoguanine DNA glycosylase (OGG1) repairs DNA by excising damaged bases (see DNA repair duties, page S21). Biochemists Zvi Livneh and Tamar Paz-Elizur, at the Weizmann Institute in Rehovot, Israel, discovered that levels of OGG1 can also be used to predict an individual's risk of developing lung cancer.

By measuring OGG1 concentration in blood samples, Livneh and Paz-Elizur found that 40% of people with lung cancer had low levels of the enzyme compared to 4% of healthy individuals. Smokers with low OGG1 activity were 5 to 10 times more likely to develop lung cancer than smokers with normal OGG1; when compared to non-smokers with normal OGG1 activity, the risk skyrocketed to 120 times more likely. The same blood test could be broadened to other cancers. For example, smokers with lower OGG1 activity are 70 times more likely to develop head and neck cancer than non-smokers with normal enzyme activity<sup>4</sup>.

➔ **NATURE.COM**  
read Missing the  
mark in this week's  
*Nature*  
[go.nature.com/4Ejeum](http://go.nature.com/4Ejeum)

OGG1 is only one of an unknown number of DNA repair enzymes; low levels of any of them could be associated with cancer development. Livneh and Paz-Elizur have expanded their research to include two additional DNA repair enzymes — AAG and APE1 — to cover people with “different risk factors to develop a certain cancer”, says Livneh. A study is underway to access their performance and results are expected in mid-2011.

It is unlikely that any single test, however many markers included, will be sufficient to gauge the risk of cancer development. “We have an additional ongoing study which explores a two-stage protocol for lung cancer prevention,” says Livneh. The first stage involves Livneh and Paz-Elizur's DNA repair biomarkers plus five biomarkers developed by other groups. These biomarkers measure: alteration in gene expression; levels of DAP kinase, an enzyme involved in programmed cell death; antibodies to mutant p53, a sign that a cell's tumour suppressor system is damaged; markers of inflammation; and variations in cancer-related genes. “Together these biomarkers are expected to yield a better risk assessment than one type alone,” says Livneh. Individuals identified as high risk will be tested using spiral computed tomography (CT). “For such a high-risk group, spiral CT early detection of lung cancer might be cost-effective and life saving,” adds Livneh.

In the initial stages of cancer, the body is often able to recognize abnormal cell

changes and raise a response, producing auto-antibodies. However, this response is limited, and in the later stages of cancer, the immune system becomes compromised and can no longer identify and attack cancer cells. Auto-antibodies are therefore prime candidates for biomarkers of early stage cancer.

By examining auto-antibody formation in presymptomatic individuals who later went on to develop lung cancer, Samir Hanash, at the Fred Hutchinson Cancer Center in Seattle, Washington, has identified three important antigens — annexin-1, 14-3-3 Theta and LAMR-1 — regarded by the immune system as foreign<sup>5</sup>. So far, specificity of these biomarkers is high but sensitivity lingers around 60%. The challenge for Hanash is to find additional candidate antigens that improve on the performance of this 3-antigen panel.

These figures might be improved by looking for even earlier signs of cancer. Through the Women's Health Initiative and Physician's Health Study, Hanash has access to blood samples that were collected up to eight years before a patient was diagnosed with lung cancer. In addition, he is searching for biomarkers of lung cancer in former smokers and in people who never smoked. “It turns out that most of the blood markers we have identified among smokers are also applicable to non-smokers,” says Hanash.

In spite of major investment in biomarker development over the past 15 years, the field of cancer prevention biomarkers looks woefully thin. One of the main reasons, according

to identify high-risk populations to be the preferred subjects for the trials. The second is surrogate endpoints that can provide evidence of whether a preventive drug is working — and do this in just a few years, rather than decades. The key to both is finding better biomarkers — the genes, proteins, and cellular metabolites that can be measured and associated with the development of cancer.

Patterns of these biomarkers that can be uniquely linked with one type of cancer can make it easier to estimate an individual's cancer risk. Selecting highest-risk patients for studies increases the statistical power of trials with a smaller number of participants. As a second benefit, high-risk cohorts can also shorten trials. If epidemiological studies show, for example, that a known percentage of patients carrying a certain gene will develop cancer within five years, researchers can restrict a prevention trial to those patients

and run it for just that duration. Moreover, patients and regulators are likely to be more tolerant of side effects if the targeted users have a high chance of developing cancer without intervention.

The designers of the dutasteride trial did select participants judged to be at higher risk of developing prostate cancer. However, they did so by looking for elevated levels of prostate-specific antigen (PSA) — a protein whose utility as a biomarker for prostate cancer is a matter of debate. If a fully validated biomarker for prostate cancer had existed, GSK might have been able to design a dutasteride trial that required fewer participants and could have yielded a more definitive outcome. In particular, looking at the drug's effect (or lack thereof) on the biomarker may have clarified whether the increase in detected higher-grade cancers was due to the drug or simply an artefact of the tumours becoming more easily detected owing to dutasteride's

shrinking of the prostate.

Some biomarkers may even function as the surrogate endpoints needed to shorten prevention trials. If, say, a specific group of proteins reliably increases in the blood of patients during the earliest, precancerous stages of disease, doctors could monitor those proteins rather than relying on biopsies to detect malignancy. Molecular biomarkers of potential toxicity, such as the activity of drug-metabolizing enzymes, could also help researchers monitor subjects' safety and response to drug candidates in clinical trials.

Scott Lippman, an oncologist and cancer prevention researcher at the University of Texas MD Anderson Cancer Center, has proposed fully integrating biomarkers chemoprevention development. After evaluating biomarkers in animal models, researchers would do epidemiologic studies linking the biomarkers to human cancers. They would next model the likelihood that

patients with specific biomarkers will develop cancer. Then, in a 'phase 0' step between preclinical and phase I clinical trials, researchers could test sub-therapeutic doses to assess a drug's behaviour in healthy patients without risking harm. Lippman argues that this approach could yield better decisions on whether to undertake a lengthy, and costly phase III trial — and speed the development of preventive agents. Indeed, the fact that GlaxoSmithKline skipped some of these steps might have played a role in the FDA's decision on dutasteride. The drug inhibits the enzyme that converts testosterone to the more potent 5 $\alpha$ -dihydrotestosterone. But neither molecule is yet a validated biomarker for prostate cancer.

**Erika Jonietz** is a science writer in Austin, Texas.

1. Andriole, G. L. *et al.* *N. Engl. J. Med.* **362**, 1192–1202 (2010).
2. Thompson, I. M. *et al.* *N. Engl. J. Med.* **349**, 215–224 (2003).

to Eleftherios Diamandis, a clinical biochemist at the University of Toronto, is because of poor study design with weak endpoints and little statistical rigor<sup>6</sup>. Furthermore, most research efforts have focused on biomarkers that monitor treatment. In fact, most biomarkers in clinical use are not suitable for population screening or for early diagnosis, observes Diamandis.

Diamandis claims that previous research into cancer biomarkers were looking in the wrong places. Too often efforts have focused on genetic markers, which in terms of cancer "represent 'digital information' — yes or no. This is not true for metabolomic or proteomic biomarkers, which are associated with quantitative changes", he says. But such biological

markers are delicate. "They can be influenced by sample collection and storage methods, benign diseases, and even diet and drugs," he explains. A difficulty of identifying quantitative biomarkers that are both highly sensitive and highly specific is that data analytical biases are introduced. "It is not surprising that seemingly spectacular data on new biomarkers are subsequently found to be not reproducible, and therefore unsuitable for use in clinical practice," Diamandis concludes.

George Poste, head of the Complex Adaptive Systems Initiative at Arizona State University in Tempe, agrees that biomarker research is yet to deliver on its promise for these and other reasons. Part of the problem, he says, is that until recently, most investigator-initiated

research has been too small and non-uniform to yield meaningful results. A lack of standardization in sample collection and processing, the use of cell lines instead of patient biopsies for research, and an insufficient number of patient samples are reasons for the dearth of meaningful biomarker development. Moreover, the field needs much more funding to encourage collaborative research and a 'big science' approach, says Poste. Government and industry funding must step up to the plate, he adds.

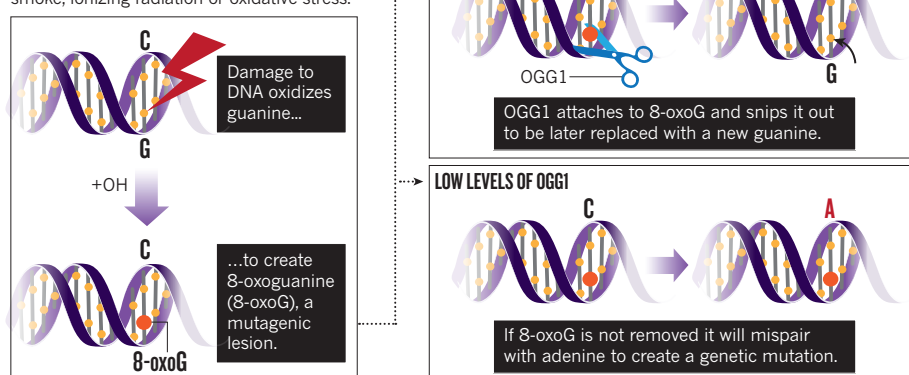
Poste cites the US National Cancer Institute's Cancer Human Biobank and the UK's Biobank as successful examples of big science and what it can do when the community invests in this research. He notes that historically the lion's share of cancer funding has gone to treatment, not prevention. "But the real issue is how can we catch cancer very early on, before it spreads," says Poste. This is the realm of biomarkers. "If we can find cancer in its earliest stages, it might be possible in the future to prevent it." ■

**Vicki Brower** is a freelance writer in New York.

1. Zhang, X. *et al.* *Physiological Genomics* **41**, 1–8 (2010).
2. Gustafson, A. *et al.* *Science Translational Medicine* **2**, 26ra25 (2010).
3. Lam, S. *et al.* *Cancer Epidemiol. Biomarkers Prev.* **15**(8): 1526–1533 (2006).
4. Paz-Elizur, T. *et al.* *Cancer Research* **66**(24), 11683–11689 (2006).
5. Qui, J. *et al.* *J. Clinical Oncology* **26**, 5060–5066 (2008).
6. Diamandis, E. J. *Natl. Cancer Inst.* **102**, 1449 (2010).

#### DNA REPAIR DUTIES

8-Oxoguanine DNA glycosylase (OGG1) removes bases that have been damaged by tobacco smoke, ionizing radiation or oxidative stress.



## DESIGNING SMARTER CANCER PREVENTION TRIALS

In April 2010, an international team of researchers from academia and drug company GlaxoSmithKline reported that dutasteride, a drug already approved for the treatment of benign prostatic hyperplasia, reduced the chances that men considered at high risk for prostate cancer would develop the disease. The four-year trial included more than 8,100 men and met the gold standard for clinical trials: it was randomized, double-blind, and placebo-controlled; it studied parallel groups at multiple medical centres; and it assessed outcomes with biopsies at two years and four years. In the end, men who took dutasteride were 23% less likely to have a positive biopsy for cancer than those on the placebo. GSK submitted this data in its application to the US Food and Drug Administration to market the drug for prostate cancer prevention



— this January, the FDA said No.

Although it is not unusual for the FDA to reject a drug application supported by apparently positive data, this case illustrates the particular challenges surrounding clinical trials for cancer prevention. When the aim is to decrease the incidence of cancer in large

populations, studies on preventive agents require large patient cohorts — sometimes approaching 20,000 participants — and take years or even decades to perform. This combination makes them especially unwieldy compared to tests with therapeutic compounds, which can much more quickly be seen to work,

or not, by testing them exclusively in people who have the disease. In cancer prevention drug trials, the usual gold standard barely rates a bronze.

Since preventives are intended for apparently healthy patients, trials require a high confidence that the anticipated anticancer benefit will outweigh any harmful side effects. In the dutasteride trial, statistical analysis showed that the decrease in cancer was driven mainly by a reduction in less serious tumours that might not even require treatment. In addition, men who took the drug were slightly more likely than those on a placebo to develop more aggressive tumours. The FDA's expert advisory panel concluded that the prevention benefits failed to outweigh this risk.

Researchers say two things are needed to decrease the length and size of prevention studies. One is

JAH CHIKWENDU/THE WASHINGTON POST VIA GETTY IMAGES

The PI3K pathway might also be used for chemoprevention. Early trials have shown that the administration of a compound that decreases PI3K activity causes regression of abnormal lesions in the airways of high-risk smokers<sup>3</sup>.

## DNA DAMAGE

As part of daily living, DNA frequently sustains damage. If not repaired, this can lead to mutations that replicate, resulting in abnormal and then cancerous growths. Certain mechanisms usually prevent this from occurring. The enzyme 8-oxoguanine DNA glycosylase (OGG1) repairs DNA by excising damaged bases (see DNA repair duties, page S21). Biochemists Zvi Livneh and Tamar Paz-Elizur, at the Weizmann Institute in Rehovot, Israel, discovered that levels of OGG1 can also be used to predict an individual's risk of developing lung cancer.

By measuring OGG1 concentration in blood samples, Livneh and Paz-Elizur found that 40% of people with lung cancer had low levels of the enzyme compared to 4% of healthy individuals. Smokers with low OGG1 activity were 5 to 10 times more likely to develop lung cancer than smokers with normal OGG1; when compared to non-smokers with normal OGG1 activity, the risk skyrocketed to 120 times more likely. The same blood test could be broadened to other cancers. For example, smokers with lower OGG1 activity are 70 times more likely to develop head and neck cancer than non-smokers with normal enzyme activity<sup>4</sup>.

➔ **NATURE.COM**  
read Missing the  
mark in this week's  
*Nature*  
[go.nature.com/4Ejeum](http://go.nature.com/4Ejeum)

OGG1 is only one of an unknown number of DNA repair enzymes; low levels of any of them could be associated with cancer development. Livneh and Paz-Elizur have expanded their research to include two additional DNA repair enzymes — AAG and APE1 — to cover people with “different risk factors to develop a certain cancer”, says Livneh. A study is underway to access their performance and results are expected in mid-2011.

It is unlikely that any single test, however many markers included, will be sufficient to gauge the risk of cancer development. “We have an additional ongoing study which explores a two-stage protocol for lung cancer prevention,” says Livneh. The first stage involves Livneh and Paz-Elizur's DNA repair biomarkers plus five biomarkers developed by other groups. These biomarkers measure: alteration in gene expression; levels of DAP kinase, an enzyme involved in programmed cell death; antibodies to mutant p53, a sign that a cell's tumour suppressor system is damaged; markers of inflammation; and variations in cancer-related genes. “Together these biomarkers are expected to yield a better risk assessment than one type alone,” says Livneh. Individuals identified as high risk will be tested using spiral computed tomography (CT). “For such a high-risk group, spiral CT early detection of lung cancer might be cost-effective and life saving,” adds Livneh.

In the initial stages of cancer, the body is often able to recognize abnormal cell

changes and raise a response, producing auto-antibodies. However, this response is limited, and in the later stages of cancer, the immune system becomes compromised and can no longer identify and attack cancer cells. Auto-antibodies are therefore prime candidates for biomarkers of early stage cancer.

By examining auto-antibody formation in presymptomatic individuals who later went on to develop lung cancer, Samir Hanash, at the Fred Hutchinson Cancer Center in Seattle, Washington, has identified three important antigens — annexin-1, 14-3-3 Theta and LAMR-1 — regarded by the immune system as foreign<sup>5</sup>. So far, specificity of these biomarkers is high but sensitivity lingers around 60%. The challenge for Hanash is to find additional candidate antigens that improve on the performance of this 3-antigen panel.

These figures might be improved by looking for even earlier signs of cancer. Through the Women's Health Initiative and Physician's Health Study, Hanash has access to blood samples that were collected up to eight years before a patient was diagnosed with lung cancer. In addition, he is searching for biomarkers of lung cancer in former smokers and in people who never smoked. “It turns out that most of the blood markers we have identified among smokers are also applicable to non-smokers,” says Hanash.

In spite of major investment in biomarker development over the past 15 years, the field of cancer prevention biomarkers looks woefully thin. One of the main reasons, according



to identify high-risk populations to be the preferred subjects for the trials. The second is surrogate endpoints that can provide evidence of whether a preventive drug is working — and do this in just a few years, rather than decades. The key to both is finding better biomarkers — the genes, proteins, and cellular metabolites that can be measured and associated with the development of cancer.

Patterns of these biomarkers that can be uniquely linked with one type of cancer can make it easier to estimate an individual's cancer risk. Selecting highest-risk patients for studies increases the statistical power of trials with a smaller number of participants. As a second benefit, high-risk cohorts can also shorten trials. If epidemiological studies show, for example, that a known percentage of patients carrying a certain gene will develop cancer within five years, researchers can restrict a prevention trial to those patients

and run it for just that duration. Moreover, patients and regulators are likely to be more tolerant of side effects if the targeted users have a high chance of developing cancer without intervention.

The designers of the dutasteride trial did select participants judged to be at higher risk of developing prostate cancer. However, they did so by looking for elevated levels of prostate-specific antigen (PSA) — a protein whose utility as a biomarker for prostate cancer is a matter of debate. If a fully validated biomarker for prostate cancer had existed, GSK might have been able to design a dutasteride trial that required fewer participants and could have yielded a more definitive outcome. In particular, looking at the drug's effect (or lack thereof) on the biomarker may have clarified whether the increase in detected higher-grade cancers was due to the drug or simply an artefact of the tumours becoming more easily detected owing to dutasteride's

shrinking of the prostate.

Some biomarkers may even function as the surrogate endpoints needed to shorten prevention trials. If, say, a specific group of proteins reliably increases in the blood of patients during the earliest, precancerous stages of disease, doctors could monitor those proteins rather than relying on biopsies to detect malignancy. Molecular biomarkers of potential toxicity, such as the activity of drug-metabolizing enzymes, could also help researchers monitor subjects' safety and response to drug candidates in clinical trials.

Scott Lippman, an oncologist and cancer prevention researcher at the University of Texas MD Anderson Cancer Center, has proposed fully integrating biomarkers chemoprevention development. After evaluating biomarkers in animal models, researchers would do epidemiologic studies linking the biomarkers to human cancers. They would next model the likelihood that

patients with specific biomarkers will develop cancer. Then, in a 'phase 0' step between preclinical and phase I clinical trials, researchers could test sub-therapeutic doses to assess a drug's behaviour in healthy patients without risking harm. Lippman argues that this approach could yield better decisions on whether to undertake a lengthy, and costly phase III trial — and speed the development of preventive agents. Indeed, the fact that GlaxoSmithKline skipped some of these steps might have played a role in the FDA's decision on dutasteride. The drug inhibits the enzyme that converts testosterone to the more potent 5 $\alpha$ -dihydrotestosterone. But neither molecule is yet a validated biomarker for prostate cancer.

**Erika Jonietz** is a science writer in Austin, Texas.

1. Andriole, G. L. *et al.* *N. Engl. J. Med.* **362**, 1192–1202 (2010).
2. Thompson, I. M. *et al.* *N. Engl. J. Med.* **349**, 215–224 (2003).

to Eleftherios Diamandis, a clinical biochemist at the University of Toronto, is because of poor study design with weak endpoints and little statistical rigor<sup>6</sup>. Furthermore, most research efforts have focused on biomarkers that monitor treatment. In fact, most biomarkers in clinical use are not suitable for population screening or for early diagnosis, observes Diamandis.

Diamandis claims that previous research into cancer biomarkers were looking in the wrong places. Too often efforts have focused on genetic markers, which in terms of cancer “represent ‘digital information’ — yes or no. This is not true for metabolomic or proteomic biomarkers, which are associated with quantitative changes”, he says. But such biological

markers are delicate. “They can be influenced by sample collection and storage methods, benign diseases, and even diet and drugs,” he explains. A difficulty of identifying quantitative biomarkers that are both highly sensitive and highly specific is that data analytical biases are introduced. “It is not surprising that seemingly spectacular data on new biomarkers are subsequently found to be not reproducible, and therefore unsuitable for use in clinical practice,” Diamandis concludes.

George Poste, head of the Complex Adaptive Systems Initiative at Arizona State University in Tempe, agrees that biomarker research is yet to deliver on its promise for these and other reasons. Part of the problem, he says, is that until recently, most investigator-initiated

research has been too small and non-uniform to yield meaningful results. A lack of standardization in sample collection and processing, the use of cell lines instead of patient biopsies for research, and an insufficient number of patient samples are reasons for the dearth of meaningful biomarker development. Moreover, the field needs much more funding to encourage collaborative research and a ‘big science’ approach, says Poste. Government and industry funding must step up to the plate, he adds.

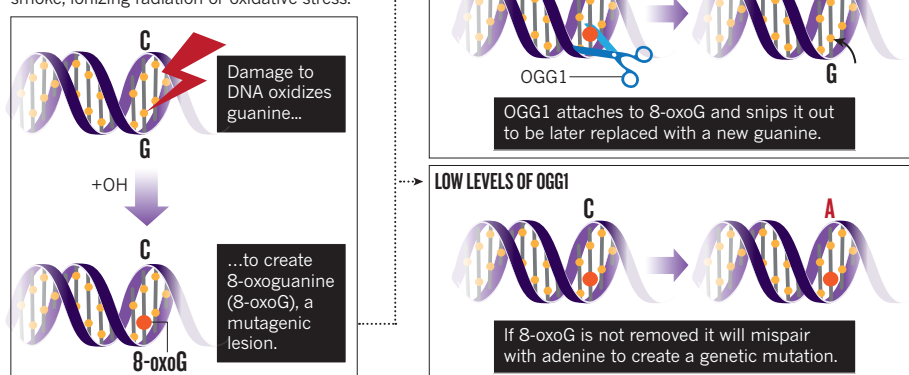
Poste cites the US National Cancer Institute's Cancer Human Biobank and the UK's Biobank as successful examples of big science and what it can do when the community invests in this research. He notes that historically the lion's share of cancer funding has gone to treatment, not prevention. “But the real issue is how can we catch cancer very early on, before it spreads,” says Poste. This is the realm of biomarkers. “If we can find cancer in its earliest stages, it might be possible in the future to prevent it.” ■

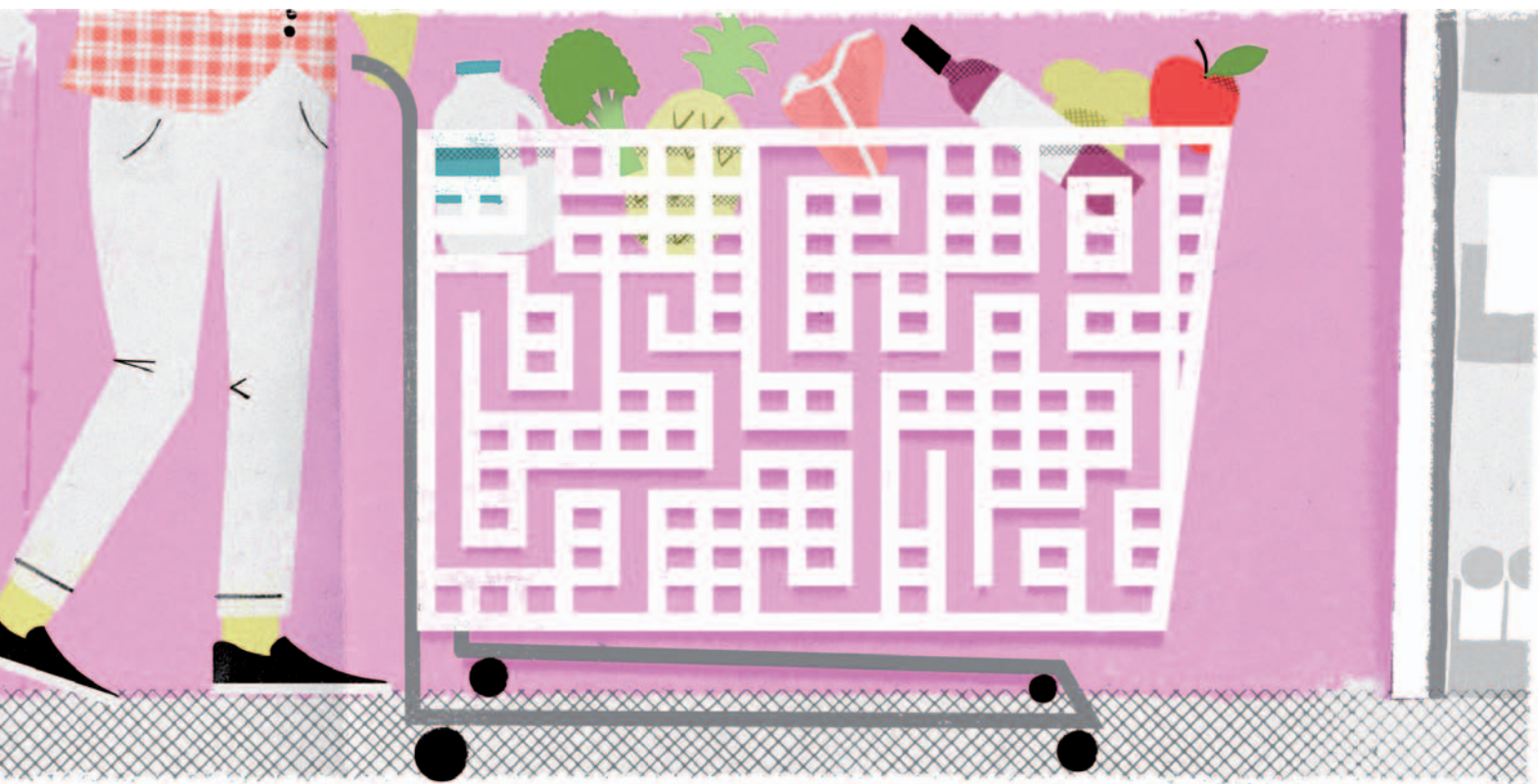
**Vicki Brower** is a freelance writer in New York.

1. Zhang, X. *et al.* *Physiological Genomics* **41**, 1–8 (2010).
2. Gustafson, A. *et al.* *Science Translational Medicine* **2**, 26ra25 (2010).
3. Lam, S. *et al.* *Cancer Epidemiol. Mark. Prev.* **15**(8): 1526–1533 (2006).
4. Paz-Elizur, T. *et al.* *Cancer Research* **66**(24), 11683–11689 (2006).
5. Qui, J. *et al.* *J. Clinical Oncology* **26**, 5060–5066 (2008).
6. Diamandis, E. J. *Natl. Cancer Inst.* **102**, 1449 (2010).

#### DNA REPAIR DUTIES

8-Oxoguanine DNA glycosylase (OGG1) removes bases that have been damaged by tobacco smoke, ionizing radiation or oxidative stress.





FOOD

# The omnivore's labyrinth

*Finding the right food to help reduce our chances of cancer can be a maze. But ongoing studies and a little inventive cooking might point us in the right direction.*

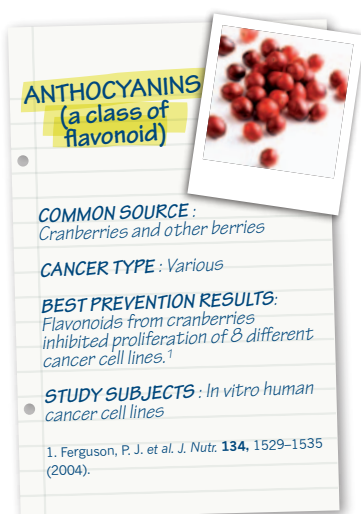
BY SARAH DEWEERDT

Twenty years ago, Paul Talalay was looking for new ways to prevent cancer, so he went grocery shopping. As a result of his trip to the supermarket, the Johns Hopkins pharmacologist and molecular scientist discovered sulforaphane, a compound contained in certain leafy vegetables. In a simple assay using mouse cells, Talalay and colleagues showed that sulforaphane dramatically boosts the activity of certain phase II enzymes, which form part of the body's cancer-fighting machinery<sup>1</sup>. Later, they demonstrated sulforaphane's capacity to prevent tumour growth in rats exposed to a carcinogen<sup>2</sup>.

The rest of us haven't been so fortunate with our anticancer shopping basket. Despite much research over the past 40 years, it's still not clear what to eat — or not eat — to help prevent cancer. Promising initial findings have often turned into statistical dead ends, leaving us culinarily confused.

In the mid-1970s, epidemiological studies suggested that people who ate more fruits and vegetables were at lower risk of several

cancers<sup>3</sup>. Such findings led to public health efforts to get people to eat more fresh produce. In both the United States and the United Kingdom, for example, official dietary recommendations pushed for everyone to consume five portions of fruit and vegetables each day.



As epidemiologists began to use large-scale, prospective studies — a more powerful type of investigation — they frequently found weak, inconsistent or no links between fruit and vegetable consumption and cancer risk. Last year, University of Oxford epidemiologist Tim Key concluded from nearly three dozen large studies and meta-analyses from the last 20 years that “at least in relatively well-nourished Westernised populations, a general increase in total fruit and vegetable intake will not have a large impact on cancer rates” (ref. 4).

This is the story for many other aspects of diet. A high-fat diet was thought to lead to an increased risk of breast cancer, until it wasn't. Early studies of dietary fibre suggested that higher consumption could decrease chances of colon cancer, but that didn't pan out either.

And yet, as Talalay found, some foods do contain anticancer compounds. These so-called phytonutrients include resveratrol in grapes, curcumin in turmeric, and many others (see recipe cards). Some of these molecules, including sulforaphane and genistein (an isoflavone found in soybeans), are now on

MATTHEW HOLLISTER



## ISOTHIOCYANATES

● **COMMON SOURCE:**  
Broccoli

**CANCER TYPE:**  
Breast

**BEST PREVENTION RESULTS:** Highest consumption (about 1 kg per month) had 60% of the risk of lowest consumption (< 350 g per month).<sup>1</sup>

● **STUDY SUBJECTS:**  
Human females

1. Ambrosone, C. B. et al. *J. Nutr.* **134**, 1134–1138 (2004).



their way to becoming pharmaceutical agents in cancer prevention (see First line of defence, page S5).

So why is figuring out the protective qualities of food so complicated? As scientists probe the complexities of diet and cancer, they are finding a host of reasons, from differences in types of vegetables to variations in human metabolism. At least seven factors underscore the difficulties.

1. *All vegetables are not created equal.* Phytonutrients are most often found in pungent vegetables (like onions and garlic), bitter ones (like mustard greens), or ones with acquired tastes (like mushrooms). There are also “a number of fruit and vegetables that are not major sources of phytonutrients,” says Johanna Lampe, an epidemiologist at the Fred Hutchinson Cancer Research Center in Seattle. And unfortunately, she says, these deficient foods include some of the most popular fruit and vegetables in the Western diet — apples and potatoes.

When Talalay analysed his trolley-load of fruit and vegetables from the supermarket for potential cancer-fighting activity, he found “a vast difference” between different types of produce, he recalls. The most potent were cruciferous vegetables (those in the cabbage family), specifically broccoli and its payload of glucoraphanin — the precursor of sulforaphane.

2. *All heads of broccoli are not created equal.* “So then we asked the question, is broccoli a good way to deliver sulforaphane?” says Talalay. “It turned out the answer was absolutely no” Different heads of broccoli varied 20-fold in their content of glucoraphanin. The specific variety, growing conditions, time of year, distance of transport and other factors all affect the concentration of phytonutrients.

Working with Johns Hopkins plant physiologist Jed Fahey, Talalay found that broccoli seeds were 10 to 100 times richer in glucoraphanin than adult plants, and certain varieties of seeds contained predictable amounts of the molecule. In fact, the most

consistent way to deliver glucoraphanin was to use three-day-old broccoli sprouts<sup>5</sup>.

Broccoli sprouts have become a powerful research tool for Talalay — and dozens of other scientists — to explore the role of sulforaphane in laboratory, animal and now human studies of cancer prevention. They also became a new food: Talalay likes to serve them at lab meetings with bagels and cream cheese.

3. *Human genomes vary.* Eating a known amount of phytonutrient doesn't guarantee that a predictable amount of the cancer-fighting molecule will enter the bloodstream. Differences here can be traced to variations in the genes involved in the digestive processes.

For example, the glutathione S-transferase M1 gene (*GSTM1*) influences the speed at which the body metabolizes sulforaphane and expels it in the urine. The faster it happens, the less beneficial the broccoli. Agricultural companies have developed several varieties of ‘super broccoli’ that are high in both glucoraphanin and related molecules to compensate for the effects of faster variants of the *GSTM1* gene.

*GSTM1* is the best studied of the genes that influence phytonutrient metabolism, but it is just one of a rapidly growing list. For example, people who carry two copies of a particular variant of the *UGT1A1* gene make about 30% to 40% less than normal of a type of phase II enzyme. One study has shown that people with this genotype derive more cancer-protecting benefit from eating cabbage- and carrot-family vegetables — possibly because phytonutrients in these foods boost *UGT1A1* activity closer to normal<sup>6</sup>.

4. *Human microbiomes vary.* Some of the genes that determine the power of phytonutrients are not human. Gut bacteria heavily metabolize the phytonutrients from soy, turning one type of isoflavones into another. So depending on their intestinal bacteria, two people who eat the same amount of soy each day might receive not only different quantities of isoflavones, but also different end-products<sup>7</sup>. Between 30% and 50% of people harbour bacteria that produce equol, which some scientists believe is one of the more beneficial forms of isoflavone; around 80% to 90% of people have bacteria that produce O-desmethylangolensin, a less active molecule.

Gut microbiota reflect a complex interplay of diet and genetics. For example, vegetarians are more likely to produce equol than non-vegetarians, and Korean-American women are more likely to produce equol than white American women.

5. *Timing is everything.* Epidemiological studies of Asian populations show that higher consumption of soy foods — tofu, miso and the like — is associated with lower breast cancer risk. Yet, soy seems to provide

little protection to people who otherwise eat a typical Western diet<sup>8</sup>.

Equol may be part of that story, but there is also growing evidence that the age at which a person starts eating soy is critical to its effect. “In Asian women, soy seems to be more protective, but the reason for that is probably that the Asian women started eating soy early in life versus most Caucasian women starting to eat it later,” explains Gertraud Maskarinec, an epidemiologist at the University of Hawaii Cancer Center.

Much of the evidence for this hypothesis comes from animal studies: rats fed soy when they are young have fewer mammary tumours later in life, but rats fed soy only as adults do not get this benefit. There are a few suggestive clues in humans too. For example, women in Asia who grew up eating a traditional, soy-rich diet and then move to the West as adults do not seem to increase their risk of breast cancer<sup>9</sup>.

6. *Some phytonutrients are difficult to access.* Many are found in small quantities in bulky foodstuffs, or only in particular types of seasonal fruit and vegetables. This means it is impractical to eat enough to noticeably reduce cancer risk. For example, many berries are rich in a group of phytonutrients called anthocyanins, which are antioxidants and may have other cancer-fighting effects as well. “The problem with berries is that they are expensive and seasonal, so availability is a limiting factor for a lot of people,” says Cathie Martin, a plant biologist at the John Innes Centre in Norwich, United Kingdom.

Martin led a team that genetically engineered a tomato (which has few natural anthocyanins) to contain roughly the same concentration of anthocyanins as blueberries. In 2008, they showed that this deep purple tomato, known as Del/Ros1N, slowed tumour progression in a cancer-prone strain of mice. The mice fed Del/Ros1N lived longer than mice fed either ordinary red tomatoes or standard laboratory fare<sup>10</sup>.

Unlike berries, tomatoes are one of the

## ● RESVERATROL

**COMMON SOURCE:**  
Red wine

**CANCER TYPE:** Glioma (brain)

**BEST PREVENTION RESULTS:** A daily dose of 40 mg/kg achieved long-term survival for 70% of subjects.<sup>1</sup>

● **STUDY SUBJECTS:** Rats harboring small subcutaneous gliomas

1. Tseng, S. H. et al. *Clin. Cancer Res.* **10**, 2190–2202 (2004).





## LYCOPENE

**COMMON SOURCE:**  
Tomatoes

**CANCER TYPE:**  
Prostate

**BEST PREVENTION RESULTS:**  
Highest consumers reduced cancer risk by 10–20%.<sup>1</sup>

**STUDY SUBJECTS:** Human males

1. Etminan, M. et al. *Cancer Epidemiol. Biomarkers Prev.* **13**, 340–345 (2004).



## CURCUMIN

**COMMON SOURCE:**  
Turmeric, curry

**CANCER TYPE:**  
Colon

### BEST PREVENTION RESULTS:

Reduced the number of precancerous tumors by up to 40%.<sup>1</sup>  
Reduced the number of precancerous tumors by an average of 60% and their size by an average of 50%.<sup>2</sup>

**STUDY SUBJECTS:** Mice, bred to be genetically susceptible to colon cancer.<sup>1</sup>  
Humans who carry a genetic mutation of susceptibility to colon cancer.<sup>2</sup>

1. Perkins, S. et al. *Cancer Epidemiol. Biomarkers Prev.* **11**, 535–540 (2002).

2. Cruz-Correa, M. et al. *Clin. Gastroenterol. Hepatol.* **4**, 1035–1038 (2006).



## GARLIC

**COMMON SOURCE:**  
Many different dishes

**CANCER TYPE:**  
Colorectal

**BEST PREVENTION RESULTS:**  
High-consumers have 30% lower risk of cancer than the lowest consumers.<sup>1</sup>

12-month consumption of garlic extract meant total size of polyps was one-third those of controls.<sup>2</sup>

**STUDY SUBJECTS:** Humans<sup>1</sup>  
Humans with precancerous colon polyps.<sup>2</sup>

1. Fleischauer, A. T. et al. *Am. J. Clin. Nutr.* **72**, 1047–1052 (2000).

2. Tanaka, S. et al. *J. Nutr.* **136**, 821S–826S (2006).



largest food crops in the world. Modern growing methods make them available all year round, and they're a popular and frequently consumed food. The anthocyanin-rich tomato could make potential anticancer foods much more readily available (though purple pizza sauce might take some getting used to).

Martin's team has also engineered a tomato strain that has a resveratrol concentration one thousand-times higher than in red wine. Resveratrol has anticancer activity in laboratory studies, but the quantities in red wine are so small that any benefit from the resveratrol would be far outweighed by the adverse effects of drinking excessive volumes of alcohol. "I'm not trying to take away anybody's red wine," Martin laughs — but she suggests that the tomato could be a more practical approach for resveratrol delivery.

7. *The whole (diet) is greater than the sum of its parts.* Sometimes, perhaps, it's not one food but the whole recipe that can be protective. That's what Thomas Kensler, an environmental health specialist, found when he began to study broccoli-sprout tea in Qidong, on the east coast of China. Kensler, based at Johns Hopkins University, expected the local people's gut bacteria to transform the tea's glucoraphanin into active sulforaphane. "There was tremendous interindividual variability," he says. "Some people were really good and some were really bad at this conversion reaction."

Talalay suggested to Kensler that adding a

bit of daikon radish to the tea might help eliminate these differences. Daikon radish contains an enzyme, myrosinase, that catalyses the glucoraphanin conversion. It was a neat trick — and one that demonstrates why it can be so difficult to isolate anticancer elements in a diet.

It is likely that these effects are not limited to experimental concoctions. Gut bacteria influence how phytonutrients are processed, but fibre intake and other aspects of the diet alter the gut ecosystem. Traditional Asian diets include not just large helpings of soy but also generous pours of green tea, which contains cancer-fighting epigallocatechins. Little is known about how diverse foods in a diet interact to affect cancer risk.

Specialized diets are far from the only way to decrease our chances of cancer, but they do add flexibility to cancer prevention strategies. Kensler began investigating sulforaphane after working for several years in China. One in ten people will develop liver cancer there during their lifetime so cancer prevention can make a real impact. Yet pharmaceuticals would be too expensive for many people to buy, and some people simply don't like taking pills. "I began to recognize that Western approaches aren't going to work around the world," Kensler recalls.

A broccoli-sprout tea was the perfect choice for this population. "Culturally it fits right in with their behavioural patterns and their own notions of how to preserve health," he says. Elsewhere, raw broccoli sprouts might be a more popular addition to diets. Some populations will be open to getting phytonutrients from a genetically engineered tomato, while other societies would recoil from designer foods. And in some cultures, purified phytonutrients in dietary supplement form might be the best approach. "Different delivery mechanisms are going to be appropriate for different

target populations," says Kensler.

But where does that leave the rest of us, as we ponder what foods to buy? For the time being we are stuck with familiar advice — eat more fruits and vegetables, and more whole grains. It can't hurt: after all, despite the healthy eating campaigns, fewer Americans today are getting their five-a-day than were a decade ago (yet another factor that makes interpreting studies of fruit-and-vegetable consumption and cancer rates more complicated).

Better guidance might be on the way. "A lot of the current recommendations are very generic — five 'helpings' of fruits and vegetables a day, but which ones are not specified," notes Martin. "I think we'll get closer to saying, 'eat the ones that are purple.'" Future dietary recommendations might also take genetics into account. We might learn how to tweak the composition of the gut microbiota, such as with probiotics, to maximize the cancer-fighting effects of foods. Plant breeders and farmers might pay more attention to phytonutrient content when they develop and sow vegetable varieties.

"Hopefully science can work together with the food industry to provide foods that are rich in components that are deemed healthy and safe," says Kensler. Perhaps then we will have clearer messages about how we can eat our way out of cancer danger. ■

**Sarah DeWeerd** is a freelance science writer based in Seattle.

1. Zhang, Y. et al. *Proc. Natl Acad. Sci. USA* **15**, 2399–2403 (1992).
2. Zhang, Y. et al. *Proc. Natl Acad. Sci. USA* **12**, 314731–314750 (1994).
3. Armstrong, B. & Doll, R. *Int. J. Cancer* **15**, 617–631 (1975).
4. Key, T. J. *BJC* **104**, 6–11 (2011).
5. Fahey, J. W. et al. *Proc. Natl Acad. Sci. USA* **94**, 10367–10372 (1997).
6. Navarro, S. L. et al. *Cancer Prev. Res.* **2**, 345–352 (2009).
7. Lampe, J. W. *Am. J. Clin. Nutr.* **89**, 1664S–1667S (2009).
8. Wu, A. H. *BJC* **98**, 9–14 (2008).
9. Maskarinec, G. & Noh, J. J. *Ethn. Dis.* **14**, 431–439 (2004).
10. Butelli, E. et al. *Nat. Biotechnol.* **26**, 1301–1308 (2008).

# Interaction-based quantum metrology showing scaling beyond the Heisenberg limit

M. Napolitano<sup>1</sup>, M. Koschorreck<sup>1</sup>, B. Dubost<sup>1,2</sup>, N. Behbood<sup>1</sup>, R. J. Sewell<sup>1</sup> & M. W. Mitchell<sup>1</sup>

Quantum metrology aims to use entanglement and other quantum resources to improve precision measurement<sup>1</sup>. An interferometer using  $N$  independent particles to measure a parameter  $\mathcal{X}$  can achieve at best the standard quantum limit of sensitivity,  $\delta\mathcal{X} \propto N^{-1/2}$ . However, using  $N$  entangled particles and exotic states<sup>2</sup>, such an interferometer<sup>3</sup> can in principle achieve the Heisenberg limit,  $\delta\mathcal{X} \propto N^{-1}$ . Recent theoretical work<sup>4–6</sup> has argued that interactions among particles may be a valuable resource for quantum metrology, allowing scaling beyond the Heisenberg limit. Specifically, a  $k$ -particle interaction will produce sensitivity  $\delta\mathcal{X} \propto N^{-k}$  with appropriate entangled states and  $\delta\mathcal{X} \propto N^{-(k-1/2)}$  even without entanglement<sup>7</sup>. Here we demonstrate ‘super-Heisenberg’ scaling of  $\delta\mathcal{X} \propto N^{-3/2}$  in a nonlinear, non-destructive<sup>8,9</sup> measurement of the magnetization<sup>10,11</sup> of an atomic ensemble<sup>12</sup>. We use fast optical nonlinearities to generate a pairwise photon–photon interaction<sup>13</sup> (corresponding to  $k = 2$ ) while preserving quantum-noise-limited performance<sup>7,14</sup>. We observe super-Heisenberg scaling over two orders of magnitude in  $N$ , limited at large numbers by higher-order nonlinear effects, in good agreement with theory<sup>13</sup>. For a measurement of limited duration, super-Heisenberg scaling allows the nonlinear measurement to overtake in sensitivity a comparable linear measurement with the same number of photons. In other situations, however, higher-order nonlinearities prevent this crossover from occurring, reflecting the subtle relationship between scaling and sensitivity in nonlinear systems. Our work shows that interparticle interactions can improve sensitivity in a quantum-limited measurement, and experimentally demonstrates a new resource for quantum metrology.

The most precise instruments are interferometric in nature, and operate according to the laws of quantum mechanics. A collection of particles, for example photons or atoms, is prepared in a superposition state, allowed to evolve under the action of a Hamiltonian containing an unknown parameter,  $\mathcal{X}$ , and measured in agreement with quantum measurement theory. The complementarity of quantum measurements<sup>15</sup> determines the ultimate sensitivity of these instruments.

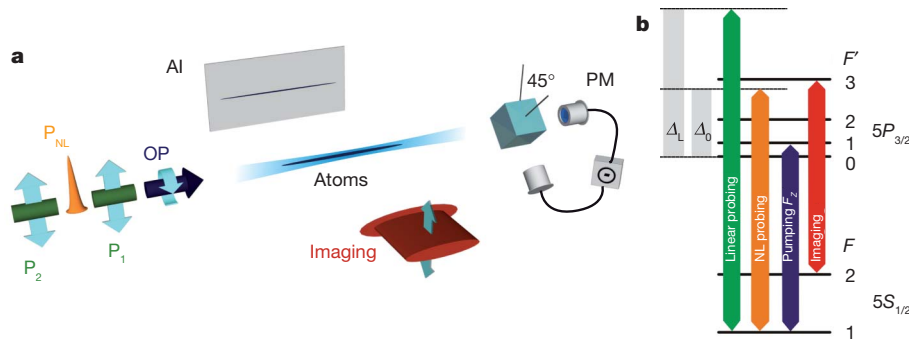
Here we describe polarization interferometry, used, for example, in optical magnetometry to detect atomic magnetization<sup>11,16,17</sup>; similar theory describes other interferometers<sup>3</sup>. A collection of  $N$  photons, with circular plus- and minus-polarization eigenstates,  $|+\rangle$  and  $|-\rangle$ , is described by single-photon Stokes operators  $\hat{S}_i = (1/2)(|+\rangle, |-\rangle)\sigma_i(|+\rangle, |-\rangle)^T$ , where  $\sigma_i$  ( $i = x, y, z$ ) are the Pauli matrices,  $\sigma_0$  is the identity and a superscript ‘T’ denotes transposition. In traditional quantum metrology, a Hamiltonian of the form  $\hat{H} = \hbar\mathcal{X} \sum_{j=1}^N \hat{S}_z^{(j)}$ , where  $\hbar$  denotes Planck’s constant divided by  $2\pi$ , uniformly and independently couples the photons to  $\mathcal{X}$ , the parameter to be measured<sup>1</sup>. If the input state consists of independent photons, the possible precision scales as  $\delta\mathcal{X} \propto N^{-1/2}$ , the shot noise or standard quantum limit (SQL). The factor of  $N^{-1/2}$  reflects the statistical averaging of independent results. In contrast, entangled states can be highly, even perfectly, correlated, giving precision limited by  $\delta\mathcal{X} \propto N^{-1}$ , the Heisenberg limit.

The above Hamiltonian is conveniently written  $\hat{H} = \hbar\mathcal{X}\hat{S}_z$ , where  $\hat{S}_i \equiv \sum_{j=1}^N \hat{S}_i^{(j)}$  is a collective variable describing the net polarization of the photons. The independence of the photons manifests itself in the linearity of this Hamiltonian. Recently, it has been shown that interactions among particles, or, equivalently, nonlinear Hamiltonians, can contribute to measurement sensitivity and give scaling beyond the Heisenberg limit<sup>4</sup>. For example, a Hamiltonian  $\hat{H} = \hbar\mathcal{X}\hat{S}_z^k$ , that is, with a  $k$ th-order nonlinearity in  $\hat{S} \equiv (\hat{S}_x, \hat{S}_y, \hat{S}_z)$ , contains  $k$ -photon interaction terms  $\hat{S}_z^{(j_1)} \otimes \hat{S}_z^{(j_2)} \otimes \dots \otimes \hat{S}_z^{(j_k)}$ . The number of such terms, and, thus, the signal strength, grows as  $N^k$ , but the quantum noise from the input states is unchanged. As a result, a sensitivity limit of  $\delta\mathcal{X} \propto N^{-k}$  applies when entanglement is used, and  $\delta\mathcal{X} \propto N^{-(k-1/2)}$  in the absence of entanglement<sup>7</sup>. For  $k \geq 2$ , this gives scaling better than the Heisenberg limit, so-called super-Heisenberg scaling<sup>7</sup>. We note that interactions and entanglement are compatible and both improve the scaling. The predicted advantage applies generally to quantum interferometry, and proposed mechanisms to produce metrologically relevant interactions include Kerr nonlinearities<sup>18</sup>, cold collisions in condensed atomic gases<sup>7</sup>, Duffing nonlinearity in nanomechanical resonators<sup>19</sup> and a two-pass effective nonlinearity with an atomic ensemble<sup>20</sup>. Topological excitations in nonlinear systems may also give advantageous scaling<sup>21</sup>.

In this Letter, we study interaction-based quantum metrology using unentangled probe particles. One challenge in demonstrating super-Heisenberg scaling is to engineer a suitable nonlinear Hamiltonian. Some nonlinearities have been shown to be intrinsically noisy<sup>14</sup> whereas others give super-Heisenberg scaling but fall short of the ideal,  $\delta\mathcal{X} \propto N^{-(k-1/2)}$ , under realistic conditions<sup>7,22</sup>. We use a cold atomic ensemble as a light–matter quantum interface<sup>12</sup> to produce quantum-noise-limited interactions, and use a Hamiltonian of the form  $\hat{H} = \hbar\mathcal{X}\hat{S}_z\hat{S}_0 = \hbar\mathcal{X}\hat{S}_zN/2$ . This Hamiltonian gives a polarization rotation that increases with photon number, without increasing quantum noise<sup>7</sup>. The experiment, shown schematically in Fig. 1, uses pulses of near-resonant light to measure the collective spin,  $\hat{\mathbf{F}}$ , of an ensemble of  $N_A \approx 10^6$  cold rubidium-87 atoms, probed on the  $5S_{1/2} \rightarrow 5P_{3/2}$   $D_2$  line. The experimental system is described in detail in refs 8, 23. The on-axis atomic magnetization,  $\langle \hat{F}_z \rangle$ , which plays the role of  $\mathcal{X}$  in this measurement, is prepared in the initial state  $\langle \hat{F}_z \rangle = N_A$  by optical pumping with resonant, circularly polarized light propagating along the trap axis,  $z$ . A weak, on-axis magnetic field is applied to preserve  $\hat{F}_z$  during the measurements.

Pulses of  $\hat{S}_x$ -polarized, but not entangled, photons pass through the ensemble and experience an optical rotation proportional to  $\langle \hat{F}_z \rangle$ . The light–atom interaction Hamiltonian  $\hat{H}_{\text{eff}} = \alpha^{(1)}\hat{F}_z\hat{S}_z + \beta^{(1)}\hat{F}_z\hat{S}_zN/2$  describes this paramagnetic Faraday rotation<sup>13</sup>. Both the linear term,  $\alpha^{(1)}\hat{F}_z\hat{S}_z$ , and the nonlinear term,  $\beta^{(1)}\hat{F}_z\hat{S}_zN/2$ , cause rotation of the plane of polarization from  $\hat{S}_x$  (vertical) towards  $\hat{S}_y$  (diagonal). Detection of  $\hat{S}_y$  then allows estimation of  $\hat{F}_z$ . As described in Supplementary Information,  $\alpha^{(1)}$  and  $\beta^{(1)}$  depend on the optical detuning,  $\Delta$ , relative to the  $F = 1 \rightarrow F' = 0$  transition; in particular,  $\alpha^{(1)}(\Delta_0) = 0$

<sup>1</sup>ICFO-Institut de Ciències Fotòniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain. <sup>2</sup>Laboratoire Matériaux et Phénomènes Quantiques, Université Paris Diderot et CNRS, UMR 7162, Bâtiment Condorcet, 75205 Paris Cedex 13, France.



**Figure 1 | Atom–light interface.** **a**, Experimental schematic: an ensemble of  $7 \times 10^5$   $^{87}\text{Rb}$  atoms, held in an optical dipole trap, is prepared in the state  $|F=1, m_F=1\rangle$  by optical pumping (OP). Linear ( $P_1, P_2$ ) and nonlinear ( $P_{\text{NL}}$ ) Faraday rotation probe pulses (in the order  $P_1, P_{\text{NL}}, P_2$ ) measure the atomic

for the specific detuning  $\Delta_0 \approx 2\pi \times 468.5$  MHz, allowing a purely non-linear estimation to be studied.

The rotation angle is  $\phi = \langle \hat{F}_z \rangle [A(\Delta) + B(\Delta)N]/2$ , where  $A \propto \alpha^{(1)}$  and  $B \propto \beta^{(1)}$  both account for the temporal pulse shape and the geometric overlap between the atomic density and the spatial mode of the probe. The shot-noise-limited uncertainty in the rotation angle, due to quantum uncertainty in the initial angle, is  $\delta\phi = N^{-1/2}$ . A contribution,  $\langle \hat{F}_z \rangle B(\Delta)\delta N/2$ , from initial number fluctuations  $\delta N = \langle N \rangle^{-1/2}$  is negligible for small rotation angles. This gives a measurement uncertainty of

$$\delta F_z = \langle \hat{F}_z \rangle \frac{\delta\phi}{\phi} = \frac{1}{A(\Delta)N^{1/2} + B(\Delta)N^{3/2}} \quad (1)$$

indicating a transition from SQL scaling,  $\delta F_z \propto N^{-1/2}$ , to super-Heisenberg scaling  $\delta F_z \propto N^{-3/2}$  with increasing  $N$ .

We use two probing regimes. The ‘linear probe’ consists of 40 1- $\mu\text{s}$  pulses (total illumination time,  $\tau_L = 40 \mu\text{s}$ ) spread over 400  $\mu\text{s}$  with detuning  $\Delta_L \gg \Delta_0$ . Together with the number of photons,  $N_L$ , used in the experiment for the linear probe, this gives  $A \gg N_L B$ , that is, linear estimation, and provides<sup>8</sup> a projection-noise-limited quantum non-demolition measurement<sup>24</sup> of  $\hat{F}_z$ , with uncertainty at the part-per-thousand level<sup>8</sup>. The ‘nonlinear probe’ consists of a single, Gaussian-shaped, high-intensity pulse with a full-duration at half-maximum of  $\tau_{\text{NL}} = 54$  ns,  $N_{\text{NL}}$  photons and a detuning  $\Delta_0$ , such that  $A \ll N_{\text{NL}} B$ . Crucially, having two probes allows us to calibrate the nonlinear measurement precisely using a highly sensitive and well-characterized independent measurement of the same sample.

We probe the same sample three times for each preparation. First we use the linear probe, which gives a precise and non-destructive measurement of  $\langle \hat{F}_z \rangle$  via the rotation angle,  $\phi_L$ . Then we use the nonlinear probe, which gives a rotation angle,  $\phi_{\text{NL}}$ , that is calibrated against the ‘true’ value (that is, with negligible error) provided by the linear probe. Finally we use a second linear probe to estimate the rotation angle  $\phi'_L$ , with which we can estimate the damage to the atomic magnetization,  $\eta \equiv 1 - \phi'_L/\phi_L$ , caused by the nonlinear probe.

The linear probe is calibrated using quantitative absorption imaging to measure  $N_A$ , and we find that  $A(\Delta_L) = 3.3(1) \times 10^{-8}$  rad per atom. The calibration of the nonlinear probe against the first linear probe is shown in Fig. 2: We repeat the above pump–probe sequence while varying  $N_A$  in the range  $1.5 \times 10^5$  to  $3.5 \times 10^5$  to generate a  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plot for a given value of  $N_{\text{NL}}$ . Because both  $\phi_L$  and  $\phi_{\text{NL}}$  are linear in  $N_A$ , we use linear regression to find the slope,  $b = d\phi_{\text{NL}}/d\phi_L = B(\Delta_0)N_{\text{NL}}/A(\Delta_L)$ , for that value of  $N_{\text{NL}}$ . The experiment is repeated for a range of different  $N_{\text{NL}}$  values.

The observed plot of  $b$  versus  $N_{\text{NL}}$ , shown in Fig. 2a, is well fitted by a simple model including saturation of the nonlinear response:

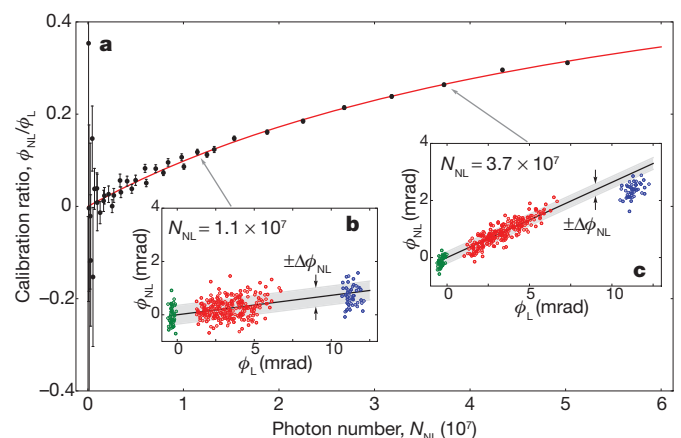
$$\frac{d\phi_{\text{NL}}}{d\phi_L} = \frac{B(\Delta_0)N_{\text{NL}}}{A(\Delta_L)} \frac{1}{1 + N_{\text{NL}}/N_{\text{NL}}^{(\text{sat})}} \quad (2)$$

magnetization, detected by a shot-noise-limited polarimeter (PM). The atom number is measured by quantitative absorption imaging (AI). **b**, Spectral positions of the pump, probe and imaging light on the  $D_2$  transition.

Here  $N_{\text{NL}}^{(\text{sat})} = 6.0(8) \times 10^7$  is a saturation parameter and the non-linear coupling strength is  $B(\Delta_0) = 3.8(2) \times 10^{-16}$  rad per atom per photon.

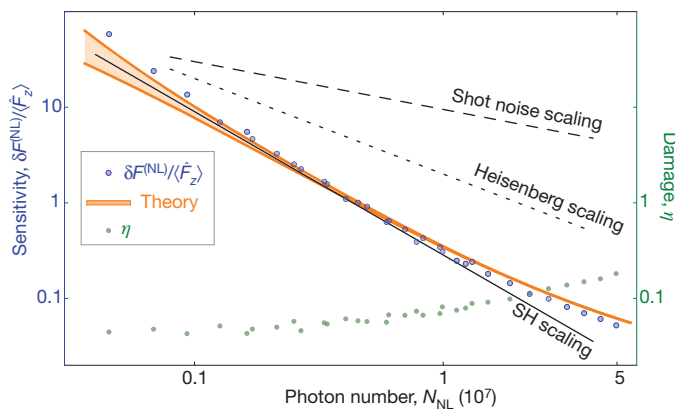
The noise in the nonlinear probe, again as a function of  $N_{\text{NL}}$ , is determined from the  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plots. As illustrated in Fig. 2b, c, the residual standard deviation of the fits indicates the observed uncertainty,  $\Delta\phi_{\text{NL}}$ , which includes the intrinsic uncertainty,  $\delta\phi_{\text{NL}}$ , and a small contribution from electronic noise. In Fig. 3, we plot the fractional sensitivity,  $\delta F_z^{(\text{NL})}/\langle \hat{F}_z \rangle$ , versus  $N_{\text{NL}}$ , calculated using equation (2) and considering the whole polarized ensemble, with  $\langle \hat{F}_z \rangle = 7 \times 10^5$  spins. In agreement with equation (1), the log–log slope indicates the scaling  $\delta F_z^{(\text{NL})} \propto N_{\text{NL}}^{-3/2}$  to within experimental uncertainties in the range  $N_{\text{NL}} = 10^6$  to  $10^7$ , and super-Heisenberg scaling, that is, steeper than  $N^{-1}$ , over two orders of magnitude ( $N_{\text{NL}} = 5 \times 10^5$  to  $5 \times 10^7$ ).

Results of numerical modelling using the Maxwell–Bloch equations to describe the nonlinear light–atom interaction are also shown in Fig. 3. Two curves are shown, for detunings  $\Delta_0 \pm (2\pi \times 200 \text{ kHz})$ , covering the combined uncertainty in  $\Delta$  due to the probe laser linewidth and inhomogeneous light shifts in the optical dipole trap. As expected from equation (1), this alters the sensitivity only at low  $N_{\text{NL}}$  values. The model is described in detail in Supplementary Information.



**Figure 2 | Calibration of nonlinear Faraday rotation.** **a**, Ratio of the nonlinear rotation,  $\phi_{\text{NL}}$ , to the linear rotation,  $\phi_L$ , versus the nonlinear probe photon number,  $N_{\text{NL}}$ . The data points and error bars indicate best-fit and standard errors from a linear regression,  $\phi_{\text{NL}} = b\phi_L + \text{const.}$ , for given values of  $N_{\text{NL}}$ . The red curve is a fit using equation (2), showing the expected nonlinear behaviour,  $\phi_{\text{NL}} \propto N_{\text{NL}}$ , with some saturation for large values of  $N_{\text{NL}}$ . **b, c**,  $\phi_L$ -vs- $\phi_{\text{NL}}$  correlation plots for two values of  $N_{\text{NL}}$ . The atom number,  $N_A$ , is varied to produce a range of  $\phi_L$  and  $\phi_{\text{NL}}$  values. Green, no atoms ( $N_A = 0$ ); red,  $1.5 \times 10^5 < N_A < 3.5 \times 10^5$ ; blue,  $N_A \approx 7 \times 10^5$ . The blue circles are shown as a check on detector saturation, and are not included in the analysis.





**Figure 3 | Super-Heisenberg scaling.** Fractional sensitivity,  $\delta F_z^{(NL)} / \langle \hat{F}_z \rangle$ , of the nonlinear probe plotted versus the number of interacting photons,  $N_{NL}$ . Blue circles indicate the measured sensitivity, orange curves show results of numerical modelling, and the black lines indicate SQL, Heisenberg-limit and super-Heisenberg (SH) scaling for reference. Scaling surpassing the Heisenberg limit,  $\propto N_{NL}^{-1}$ , is observed over two orders of magnitude. The measured damage to the magnetization,  $\eta$ , shown as green circles, confirms the non-destructive nature of the measurement. Error bars for standard errors would be smaller than the symbols and are not shown.

For photon numbers above  $N_{NL} \approx 2 \times 10^7$ , the saturation of the nonlinear rotation alters the slope. This can be understood as optical pumping of atoms into states other than  $|F=1, m_F=1\rangle$  by the nonlinear probe. The damage to the atomic magnetization,  $\eta = 1 - \phi_L / \phi_L$ , also shown in Fig. 3, remains small, confirming the non-destructive nature of the measurement. The finite damage even for small  $N_{NL}$  values is possibly due to stray light and/or magnetic fields disturbing the atoms during the 20-ms period between the two linear measurements. For large  $N$ , higher-order nonlinear effects including optical pumping limit the range of super-Heisenberg scaling.

The experimental results illustrate the subtle relationship between scaling and sensitivity in a nonlinear system. For an ideal nonlinear measurement, the improved scaling would guarantee better absolute sensitivity for sufficiently large  $N$  values. Indeed, when the measurement bandwidth is taken into account, the nonlinear probe overtakes the linear one at  $N = 3.2 \times 10^6$ , where both achieve a sensitivity of  $1.1 \times 10^2$  spins  $\text{Hz}^{-1/2}$ . As a consequence, the nonlinear technique performs better in fast measurements. In contrast, when measurement time is not a limited resource, the comparison can be made on a 'sensitivity-per-measurement' basis and the ideal crossover point, of  $3.2 \times 10^3$  spins at  $N = 8.7 \times 10^7$ , is never actually reached, owing to the higher-order nonlinearities. Evidently, super-Heisenberg scaling allows but does not guarantee enhanced sensitivity: for the nonlinear technique to overtake the linear, it is also necessary that the scaling extend to large enough values of  $N$ . This example shows also that resource constraints dramatically influence the comparison between the linear and nonlinear techniques. See also Supplementary Information.

We have experimentally realized a system designed to achieve metrological sensitivity beyond the Heisenberg limit,  $\delta \mathcal{A} \propto N^{-1}$ , using metrologically relevant interactions among particles. To generate pairwise photon-photon interactions, we use fast, nonlinear optical effects in a cold atomic ensemble and measure the ensemble magnetization,  $\langle \hat{F}_z \rangle$ , with super-Heisenberg sensitivity  $\delta F_z \propto N^{-3/2}$ . To quantify the photon-photon interaction and the sensitivity rigorously, we calibrate against a precise, non-destructive, linear measurement of the same atomic quantity<sup>8</sup>, demonstrate quantum-noise-limited performance of the optical instrumentation and place an upper limit on systematic, that is, non-atomic, nonlinearities at the level of a few per cent. The experiment demonstrates the use of interparticle interactions as a new resource for quantum metrology. Although possible applications to precision measurement will require detailed study, our experiment

shows that interactions can produce super-Heisenberg scaling and improved precision in a quantum-limited measurement.

## METHODS SUMMARY

**Linear and nonlinear probe light.** The probe beam is aligned on the axis of the trap with a waist of 20  $\mu\text{m}$ , chosen to match the radial dimension of the cloud. In the linear probing regime, we use a train of 40 1- $\mu\text{s}$  pulses, with a repetition rate of 100 kHz, each containing  $3 \times 10^6$  photons detuned by +1.5 GHz from the  $F=1 \rightarrow F'=0$  transition. The maximum intensity is  $0.1 \text{ W cm}^{-2}$ . The signals are summed and can be considered a single, modulated pulse.

The nonlinear probe consists of a single, Gaussian-shaped pulse with a full-duration at half-maximum of 54 ns. The maximum intensity of the nonlinear probe is  $7 \text{ W cm}^{-2}$  for a pulse with  $10^7$  photons. Theory predicts that  $\alpha^{(1)} = 0$  at a detuning of  $\Delta = 2\pi \times 462 \text{ MHz}$  in free space. This is modified by trap-induced light shifts, and we use the empirical value  $\Delta_0 = 2\pi \times 468.5 \text{ MHz}$ , which gives zero rotation at low probe intensity.

**Instrumental noise.** The instrumental noise is quantified by measuring  $\text{var}(\hat{S}_y)$  versus input photon number  $N$  (that is,  $N_L$  or  $N_{NL}$ ), in the absence of atoms, to find contributions from electronic noise ( $V^{(el)} \propto N^0$ ), shot noise ( $N^1$ ) and technical noise ( $\propto N^2$ ), as described in Supplementary Information. We find that the contributions from electronic noise to the linear ( $V_L^{(el)}$ ) and nonlinear ( $V_{NL}^{(el)}$ ) probes are  $3 \times 10^5$  and  $4 \times 10^5$  per pulse, respectively, and that the technical noise is negligible. The instrumentation is thus shot-noise-limited over the full range of  $N$  used in the experiment. The intrinsic rotation uncertainty of the nonlinear probe,  $\delta\phi_{NL}$ , is calculated from the measured  $\Delta\phi_{NL}$  as  $(\delta\phi_{NL})^2 = (\Delta\phi_{NL})^2 - V_{NL}^{(el)}$ . The correction is at most 5%.

**Instrumental linearity.** The linearity of the experimental system and analysis is verified by using a wave plate in place of the atoms to produce a linear rotation equal to the largest observed nonlinear rotation. Over the full range of photon numbers used in the experiment, the detected rotation angle is constant to within 5%, and SQL scaling is observed.

Received 31 July; accepted 15 December 2010.

- Giovannetti, V., Lloyd, S. & Maccone, L. Quantum metrology. *Phys. Rev. Lett.* **96**, 010401 (2006).
- Mitchell, M. W., Lundeen, J. S. & Steinberg, A. M. Super-resolving phase measurements with a multiphoton entangled state. *Nature* **429**, 161–164 (2004).
- Lee, H., Kok, P. & Dowling, J. P. A quantum Rosetta stone for interferometry. *J. Mod. Opt.* **49**, 2325–2338 (2002).
- Boixo, S., Flammaria, S. T., Caves, C. M. & Geremia, J. Generalized limits for single-parameter quantum estimation. *Phys. Rev. Lett.* **98**, 090401 (2007).
- Choi, S. & Sundaram, B. Bose-Einstein condensate as a nonlinear Ramsey interferometer operating beyond the Heisenberg limit. *Phys. Rev. A* **77**, 053613 (2008).
- Roy, S. M. & Braunstein, S. L. Exponentially enhanced quantum metrology. *Phys. Rev. Lett.* **100**, 220501 (2008).
- Boixo, S. *et al.* Quantum metrology: dynamics versus entanglement. *Phys. Rev. Lett.* **101**, 040403 (2008).
- Koschorreck, M., Napolitano, M., Dubost, B. & Mitchell, M. W. Sub-projection-noise sensitivity in broadband atomic magnetometry. *Phys. Rev. Lett.* **104**, 093602 (2010).
- Koschorreck, M., Napolitano, M., Dubost, B. & Mitchell, M. W. Quantum nondemolition measurement of large-spin ensembles by dynamical decoupling. *Phys. Rev. Lett.* **105**, 093602 (2010).
- Kominis, I., Kornack, T., Allred, J. & Romalis, M. A subfemtotesla multichannel atomic magnetometer. *Nature* **422**, 596–599 (2003).
- Budker, D. & Romalis, M. Optical magnetometry. *Nature Phys.* **3**, 227–234 (2007).
- Hammerer, K., Sørensen, A. S. & Polzik, E. S. Quantum interface between light and atomic ensembles. *Rev. Mod. Phys.* **82**, 1041–1093 (2010).
- Napolitano, M. & Mitchell, M. W. Nonlinear metrology with a quantum interface. *N. J. Phys.* **12**, 093016 (2010).
- Fleischhauer, M., Matsuoka, A. B. & Scully, M. O. Quantum limit of optical magnetometry in the presence of ac Stark shifts. *Phys. Rev. A* **62**, 013808 (2000).
- Scully, M. O., Englert, B. G. & Walther, H. Quantum optical tests of complementarity. *Nature* **351**, 111–116 (1991).
- Wasilewski, W. *et al.* Quantum noise limited and entanglement-assisted magnetometry. *Phys. Rev. Lett.* **104**, 133601 (2010).
- Wolffgramm, F. *et al.* Squeezed-light optical magnetometry. *Phys. Rev. Lett.* **105**, 053601 (2010).
- Beltrán, J. & Luis, A. Breaking the Heisenberg limit with inefficient detectors. *Phys. Rev. A* **72**, 045801 (2005).
- Woolley, M. J., Milburn, G. J. & Caves, C. M. Nonlinear quantum metrology using coupled nanomechanical resonators. *N. J. Phys.* **10**, 125018 (2008).
- Chase, B. A., Baragiola, B. Q., Partner, H. L., Black, B. D. & Geremia, J. M. Magnetometry via a double-pass continuous quantum measurement of atomic spin. *Phys. Rev. A* **79**, 062107 (2009).
- Negretti, A., Henkel, C. & Mølmer, K. Quantum-limited position measurements of a dark matter-wave soliton. *Phys. Rev. A* **77**, 043606 (2008).

22. Boixo, S. *et al.* Quantum-limited metrology and Bose-Einstein condensates. *Phys. Rev. A* **80**, 032103 (2009).
23. Kubasik, M. *et al.* Polarization-based light-atom quantum interface with an all-optical trap. *Phys. Rev. A* **79**, 043815 (2009).
24. Braginskii, V. B. & Vorontsov, Y. I. Quantum-mechanical limitations in macroscopic experiments and modern experimental technique. *Sov. Phys. Usp.* **17**, 644 (1975).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank I. H. Deutsch and F. Illuminati for comments. We thank C. M. Caves and A. D. Codorníu for inspiration. This work was supported by the Spanish

Ministry of Science and Innovation through the Consolider-Ingenio 2010 project QOIT, the Ingenio-Explora project OCHO (ref. FIS2009-07676-E/FIS) and project ILUMA (ref. FIS2008-01051), by the Marie-Curie RTN EMALI, and by Fundacio CELLEX Barcelona.

**Author Contributions** All authors contributed equally to the work presented in this paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M.N. ([mario.napolitano@icfo.es](mailto:mario.napolitano@icfo.es)).

# A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression

Kevin C. Wang<sup>1,2</sup>, Yul W. Yang<sup>1\*</sup>, Bo Liu<sup>3\*</sup>, Amartya Sanyal<sup>4</sup>, Ryan Corces-Zimmerman<sup>1</sup>, Yong Chen<sup>5</sup>, Bryan R. Lajoie<sup>4</sup>, Angeline Protacio<sup>1</sup>, Ryan A. Flynn<sup>1</sup>, Rajnish A. Gupta<sup>1</sup>, Joanna Wysocka<sup>6</sup>, Ming Lei<sup>5</sup>, Job Dekker<sup>4</sup>, Jill A. Helms<sup>3</sup> & Howard Y. Chang<sup>1</sup>

The genome is extensively transcribed into long intergenic non-coding RNAs (lincRNAs), many of which are implicated in gene silencing<sup>1,2</sup>. Potential roles of lincRNAs in gene activation are much less understood<sup>3–5</sup>. Development and homeostasis require coordinate regulation of neighbouring genes through a process termed locus control<sup>6</sup>. Some locus control elements and enhancers transcribe lincRNAs<sup>7–10</sup>, hinting at possible roles in long-range control. In vertebrates, 39 *Hox* genes, encoding homeodomain transcription factors critical for positional identity, are clustered in four chromosomal loci; the *Hox* genes are expressed in nested anterior-posterior and proximal-distal patterns colinear with their genomic position from 3' to 5' of the cluster<sup>11</sup>. Here we identify *HOTTIP*, a lincRNA transcribed from the 5' tip of the *HOXA* locus that coordinates the activation of several 5' *HOXA* genes *in vivo*. Chromosomal looping brings *HOTTIP* into close proximity to its target genes. *HOTTIP* RNA binds the adaptor protein WDR5 directly and targets WDR5/MLL complexes across *HOXA*, driving histone H3 lysine 4 trimethylation and gene transcription. Induced proximity is necessary and sufficient for *HOTTIP* RNA activation of its target genes. Thus, by serving as key intermediates that transmit information from higher order chromosomal looping into chromatin modifications, lincRNAs may organize chromatin domains to coordinate long-range gene activation.

We examined chromosome structure and histone modifications in human primary fibroblasts derived from several anatomic sites<sup>12</sup>, and found distinctive differences in the *HOXA* locus. High throughput chromosome conformation capture (5C)<sup>13</sup> across *HOXA* revealed that its higher order structure is dependent on positional identity. In anatomically distal cells (for example, foreskin and foot fibroblasts), we detected abundant chromatin interactions within the transcriptionally active 5' *HOXA* locus (with reference to the directions of transcription of constituent *Hox* genes), pointing to a compact and looped conformation. In contrast, no long-range chromatin interactions are detected within the transcriptionally silent 3' *HOXA* which seems largely linear (Fig. 1a). Strikingly, anatomically proximal cells (for example, lung fibroblasts) have the diametrically opposite pattern. The ON and OFF states of *Hox* and other key developmental genes are maintained by the MLL/Trithorax (Trx) and polycomb group (PcG) proteins, which mediate trimethylation of histone H3 lysine 4 (H3K4me3) to activate genes or lysine 27 (H3K27me3) to repress genes<sup>14</sup>. The portions of *HOXA* in tight physical interaction are marked by broad domains of H3K4me3, whereas H3K27me3 marks the physically extended and transcriptional silent regions (Fig. 1a).

On the very 5' and 3' edges of the two respective interaction clusters are two lincRNA loci that exhibit distinct chromatin modifications. The 3' element has been previously identified as the myelopoiesis-associated lincRNA *HOTAIRM1* (ref. 15). The 5' element, for which

we suggest the name *HOTTIP* for 'HOXA transcript at the distal tip', exhibits bivalent H3K4me3 and H3K27me3, a histone modification pattern associated with poised regulatory sequences<sup>16</sup>. Comparison with RNA polymerase II occupancy and RNA expression showed that the bivalent H3K4me3 and H3K27me3 modifications on *HOTTIP* gene do not require *HOTTIP* transcription, but transcription of *HOTTIP* is associated with increased H3K4me3 and decreased H3K27me3 (Fig. 1a, left). Chromatin immunoprecipitation (ChIP) analysis confirmed that the *HOTTIP* gene is occupied by both polycomb repressive complex 2 (PRC2) and MLL complex, consistent with the bivalent histone marks (Supplementary Fig. 1a).

*HOTTIP* transcription yields a 3,764-nucleotide, spliced and polyadenylated lincRNA that initiates ~330 bases upstream of *HOXA13*. Only the strand antisense to *HOXA* genes is transcribed (Supplementary Fig. 1b). Genes near the 5' end of each *HOX* cluster tend to be expressed in more posterior and/or distal anatomical locations. Consistent with its genomic location 5' to *HOXA13*, *HOTTIP* is expressed in distal and/or posterior anatomic sites (Fig. 1b). *In situ* hybridization of developing mouse and chick embryos confirmed that *HOTTIP* is expressed in posterior and distal sites *in vivo*, indicating a conserved expression pattern from development to adulthood (Fig. 1c and Supplementary Fig. 1c). Even in distal cells where *HOTTIP* is expressed, its RNA level is very low and estimated to be ~0.3 copies per cell (Supplementary Fig. 2).

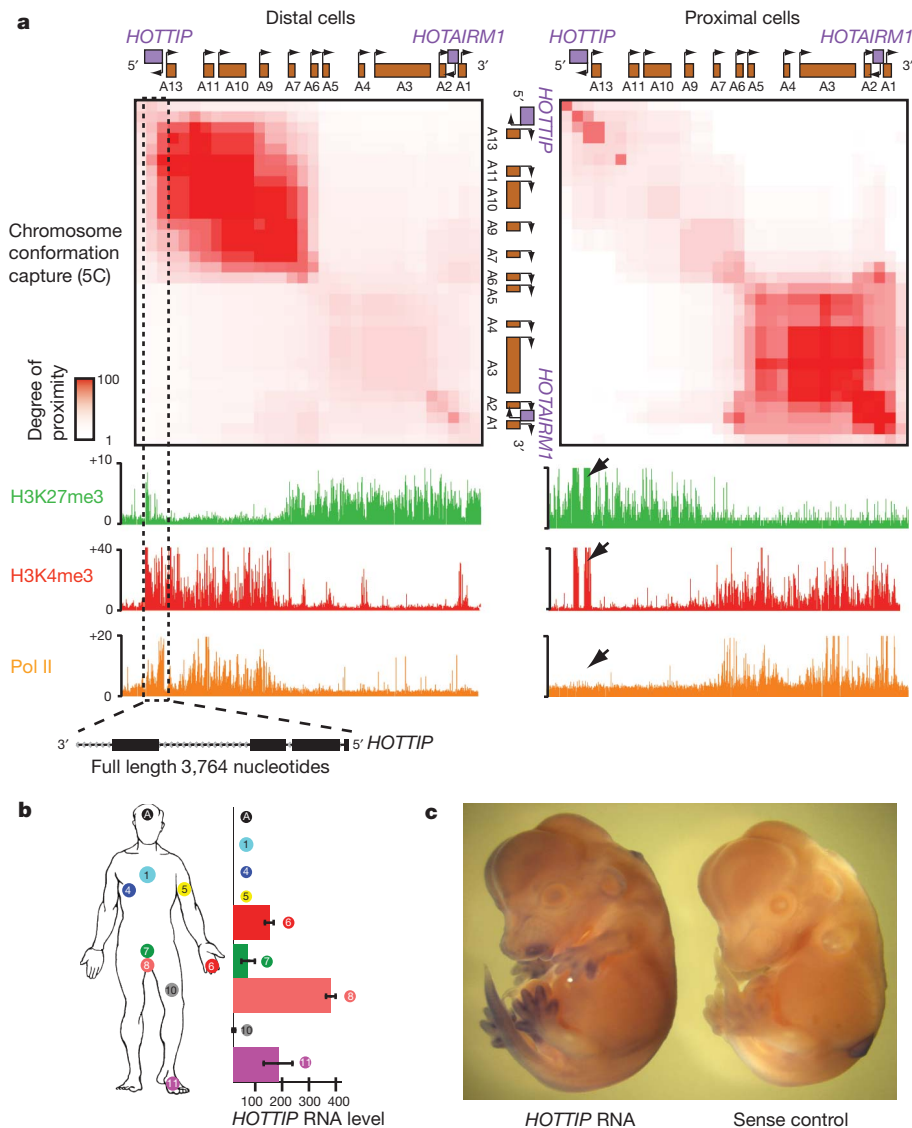
We employed small interfering RNAs (siRNAs) to knock down *HOTTIP* RNA in fibroblasts from a distal anatomic site (foreskin), and examined expression of 5' *HOXA* genes by quantitative reverse transcription PCR. Notably, *HOTTIP* RNA knockdown abrogated expression of distal *HOXA* genes across 40 kilobases with a trend dependent on the distance to *HOTTIP*. The strongest blockade was observed for *HOXA13* and *HOXA11*, with progressively less severe effects on *HOXA10*, *HOXA9* and *HOXA7* (Fig. 2a). The effect on gene transcription appeared to be unidirectional, as there were no appreciable changes in the levels of *EVX1*, located ~40 kilobases 5' of the *HOXA* cluster (data not shown). *HOTTIP* knockdown did not affect expression of the highly homologous *HOXD* genes, other control genes, nor induce antisense transcription at its own locus (Fig. 2b, Supplementary Fig. 3a). Several independent siRNAs targeting *HOTTIP* yielded similar results (Supplementary Fig. 3b). These results indicate that *HOTTIP* RNA is necessary to coordinate activation of 5' *HOXA* genes.

We next addressed the function of *HOTTIP* RNA *in vivo* in the developing chick limb bud (Fig. 2c). Whereas prior genetic studies of noncoding RNAs (ncRNAs) involved deletion or insertion into the gene locus<sup>17</sup>, we wished to distinguish the functions of *HOTTIP* RNA from its corresponding DNA element. *HOTTIP* gene can nucleate H3K4 and H3K27 methylation independent of transcription (Fig. 1a),

<sup>1</sup>Howard Hughes Medical Institute, Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>2</sup>Department of Dermatology, University of California San Francisco (UCSF), San Francisco, California 94115, USA. <sup>3</sup>Department of Surgery, Stanford University School of Medicine, Stanford, California 94305, USA. <sup>4</sup>Program in Gene Function and Expression, Dept. of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. <sup>5</sup>Howard Hughes Medical Institute, Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. <sup>6</sup>Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305.

\*These authors contributed equally to this work.





**Figure 1 | *HOTTIP* is a lincRNA transcribed in distal anatomic sites.**

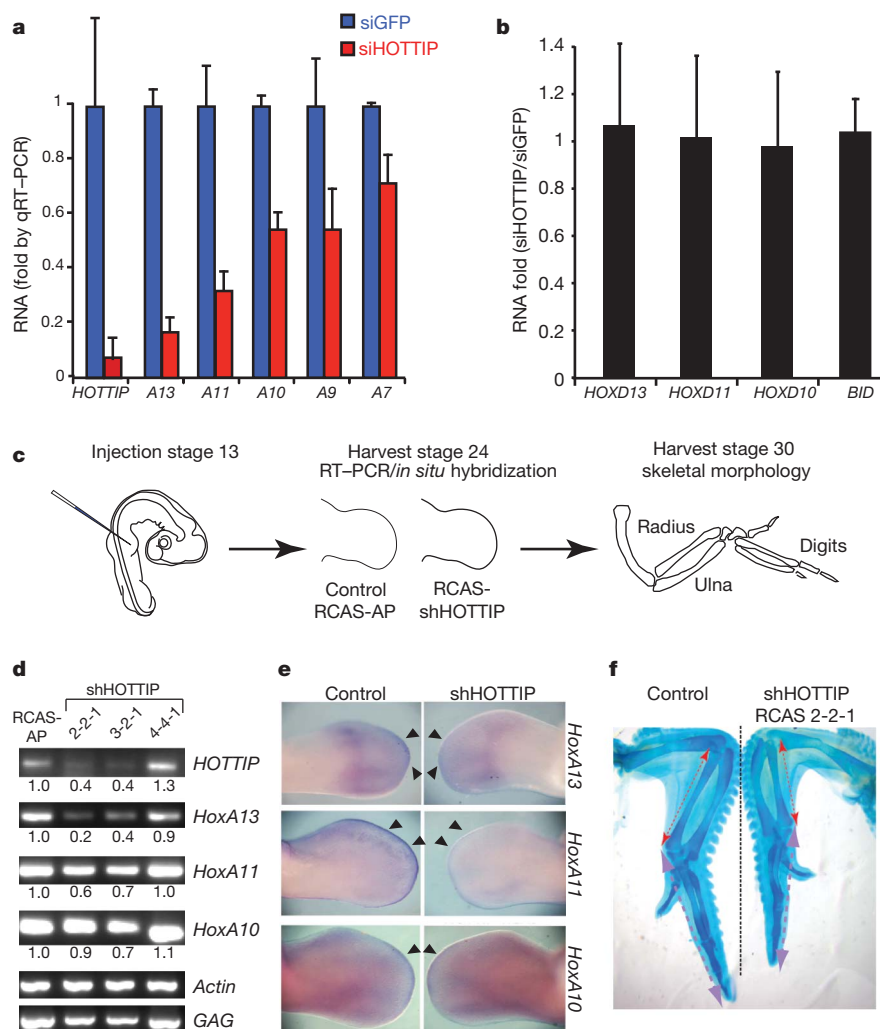
**a**, Chromatin state map of distal versus proximal cells. Top panels, chromosome conformation capture-carbon copy (5C) analysis of distal (foreskin) and proximal (lung) human fibroblasts. Heat map representations (generated by my5C, ref. 30) of 5C data (bin size 30 kb, step size 3 kb) for *HOXA* in foreskin and lung fibroblasts. Red intensity of each pixel indicates relative interaction between the two points on the genomic coordinates. The diagonal represents frequent *cis* interactions between regions located in close proximity

along the linear genome. 5C signals that are away from the diagonal represent long-range looping interactions. Bottom panels, chromatin occupancy across *HOXA*. x-axis is genomic coordinate; y-axis depicts occupancy of the indicated histone marks or protein (ChIP/input). Box and arrows highlight chromatin states of *HOTTIP* gene. **b**, *HOTTIP* RNA expression in primary human fibroblasts from 11 anatomic sites. Means  $\pm$  s.d. are shown ( $n = 2$ ). **c**, *In situ* hybridization of *HOTTIP* RNA in E13.5 mouse embryo.

and the precise genomic distance between upstream enhancer elements and *Hox* genes is critical for their proper colinear activation<sup>17</sup>. Therefore, we used RNA interference (RNAi) in chick embryos, where replication-competent retroviruses can deliver short-hairpin RNAs (shRNAs) with high penetrance and precise spatiotemporal control<sup>18</sup> (Supplementary Fig. 4). In the limb bud, 5' *HoxA* genes are transcribed in a nested pattern along the proximal-distal axis<sup>19</sup>. In this tissue, *HoxA* function is highly redundant with that of the *HoxD* locus, which allowed us to assess altered *HoxA* expression patterns without major changes in anatomic landmarks<sup>20</sup>. We injected retroviruses carrying shRNAs against chick *HOTTIP* into upper limb buds of stage 13 chicks; RT-PCR and *in situ* hybridization were performed on both control and knockdown samples after 2–4 days. Knockdown of *HOTTIP* RNA by two independent shRNAs in limb buds decreased expression of *HoxA13*, *HoxA11* and *HoxA10*—again with a graded impact depending on genomic proximity to *HOTTIP* gene. Vector control or an shRNA that fails to deplete *HOTTIP* RNA had little effect on *Hox* gene

expression (Fig. 2d). *In situ* hybridization on whole embryos (Fig. 2e) and sections (Supplementary Fig. 5) revealed that *HOTTIP* RNA most strongly affects *HoxA* gene expression at the distal edge of the developing limb bud, where the 5' *HoxA* genes are most strongly expressed. By stage 36, limbs depleted of *HOTTIP* RNA showed notable shortening and bending of distal bony elements, including the radius, ulna and third digit (~20% length reduction for each compared to contralateral and stage-matched limbs treated with control virus,  $P < 0.05$ , Student's *t*-test, Fig. 2f). This phenotype resembled some of the defects in mice lacking *HoxA11* and *HoxA13* (refs 21–23). Together, these data indicate that *HOTTIP* RNA controls activation of distal *Hox* genes *in vivo*.

The broad impact of *HOTTIP* RNA on gene activation across the *HOXA* locus is reminiscent of the broad domains of chromatin modifications demarcating active and silent chromosomal domains<sup>12</sup>. 5C analysis of control and *HOTTIP*-depleted cells showed little change in higher order chromosomal structure, indicating that the chromosomal looping is pre-configured and upstream of gene expression



**Figure 2 | *HOTTIP* is required for coordinate activation of 5' *HOXA* genes.**

**a, b**, Knockdown of *HOTTIP* RNA abrogates expression of 5' *HOXA* genes in fore skin fibroblasts (**a**), but not *HOXD* or *BID* genes (**b**). Means  $\pm$  s.d. are shown ( $n = 3$ ). GFP, green fluorescent protein. **c**, Schematic of chick RNAi experiment. **d**, *HOTTIP* RNA is required for 5' *HoxA* gene expression *in vivo*. RT-PCR of the indicated genes from control or *HOTTIP*-depleted distal limb bud is shown; quantification and normalization by *Actin* signal is shown below

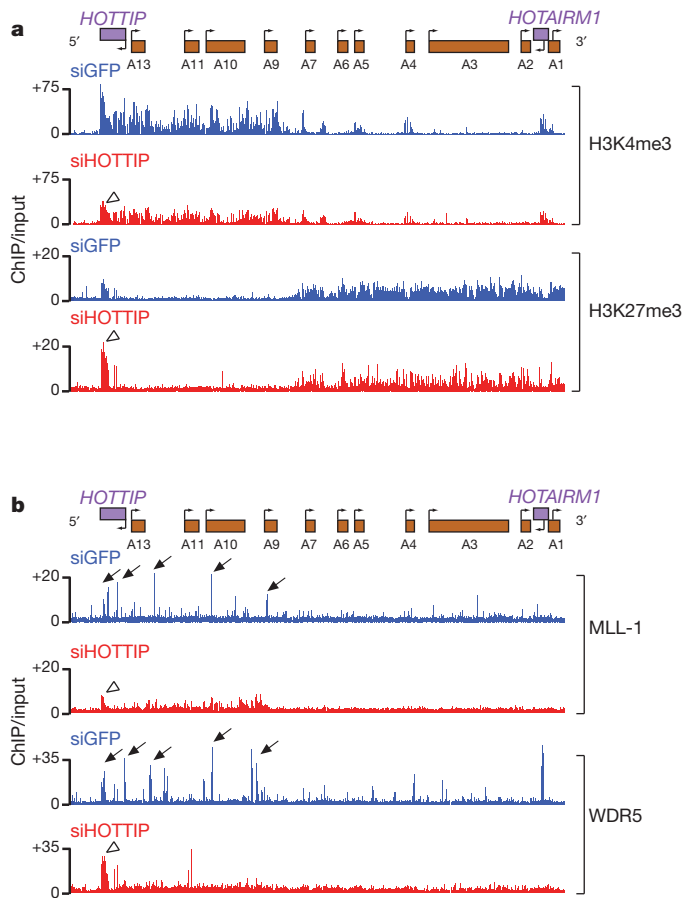
each band. *GAG* signal confirms successful retroviral transduction in all cases. **e**, *In situ* hybridization of 5' *HoxA* genes in chick limb buds. Arrowheads highlight distal domains of high *HoxA* gene expression that are affected by *HOTTIP* knockdown. **f**, Shortening of distal bony elements in *HOTTIP*-depleted forelimbs. Alcan blue staining highlights the skeletal elements. Red and purple lines highlight radius and 3rd digit lengths, respectively.

(Supplementary Fig. 6a). In contrast, *HOTTIP* RNA knockdown led to broad loss of H3K4me3 and H3K4me2 across the *HOXA* locus, most prominently over 5' *HOXA* and *HOTTIP* gene itself (Fig. 3a, Supplementary Figs 6b and 7). *HOTTIP* RNA knockdown also increased H3K27me3 focally over *HOTTIP* gene, but had little impact on H3K27me3 across *HOXA*. These results indicate that *HOTTIP* RNA is required for maintenance of H3K4me3 across the *HOXA*. These findings also imply that loss of 5' *HOXA* gene transcription upon *HOTTIP* RNA knockdown is likely to be due to loss of H3K4me3 (or other changes) rather than ectopic spread of H3K27me3.

H3K4 methylation of the *HOX* loci is carried out by the MLL family of complexes<sup>24</sup>. In mammals, at least six MLL family members of SET-domain-containing lysine methyltransferases interact with a core complex of WDR5, ASH2L, RBBP5, as well as with other proteins, for substrate recognition and genomic targeting<sup>24</sup>. Genetic analyses indicate that MLL1 and 2 are most essential for *HOX* gene expression in fibroblasts<sup>25</sup>, and MLL1 in particular is recruited to promoters of *HOX* genes to maintain their activation states<sup>26</sup>. In distally-derived human fibroblasts, MLL1 and WDR5 densely occupied extended region of the 5' *HOXA* cluster, coincident with the H3K4me3 domain, with specific

'peaks' of occupancy near the transcriptional start sites (TSS) of multiple 5' *HOXA* genes (Fig. 3b). Strikingly, *HOTTIP* RNA knockdown abrogated the peaks of MLL1 and WDR5 occupancy near TSS, resulting in diffuse and less intense binding of MLL1 and WDR5 across *HOXA* cluster, most prominently over the 5' *HOXA* domain. *HOTTIP* RNA knockdown also led to increased accumulation of MLL1 and WDR5 on *HOTTIP* gene itself (Supplementary Fig. 8). Thus, *HOTTIP* RNA seems critical for maintaining a specific pattern of MLL complex occupancy across the *HOXA* locus to facilitate H3K4me3 and active transcription.

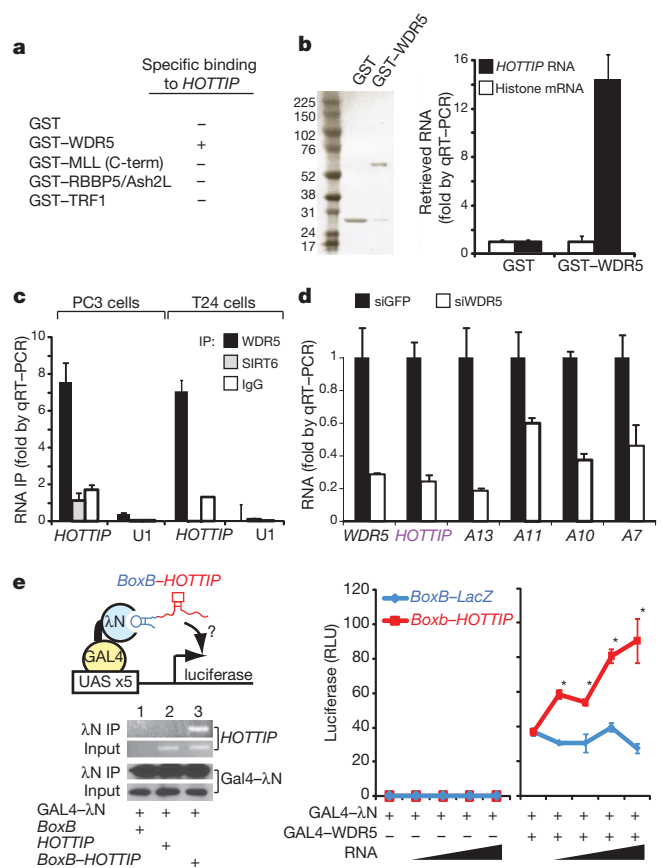
To define the molecular link between *HOTTIP* RNA and MLL complex, we reasoned that *HOTTIP* RNA may physically interact with one or more subunits of the MLL complex. Purified, *in-vitro*-transcribed, full-length *HOTTIP* RNA bound specifically to recombinant glutathione-S-transferase-conjugated WDR5 (GST-WDR5), but not to GST, RBBP5, ASH2L, or the telomeric protein TRF1 (also known as TERF1; Fig. 4a, b). The C terminus of MLL1, containing the SET domain, bound non-specifically to all RNAs, consistent with previous studies<sup>27</sup>. Immunoprecipitation of endogenous WDR5 from two different cell lines each specifically retrieved endogenous *HOTTIP* RNA (Fig. 4c), indicating that WDR5 and *HOTTIP* RNA interact in living



**Figure 3 | *HOTTIP* RNA is required for the active chromatin state of 5' *HOXA* cluster.** **a**, Knockdown of *HOTTIP* RNA broadly decreases H3K4me3 across 5' *HOXA* locus but focally affects H3K27me3 at *HOTTIP* gene. Display is as in Fig. 1a. **b**, Knockdown of *HOTTIP* RNA abrogates peaks of MLL1 and WDR5 occupancy near TSSs of 5' *HOXA* genes and leads to accumulation of these proteins at *HOTTIP* gene itself. Arrows highlight peaks of MLL1 and WDR5 occupancy; open arrowheads highlight chromatin state of *HOTTIP* gene upon *HOTTIP* RNA knockdown.

cells. Immunoprecipitation of an epitope-tagged WDR5 from a stable cell line that previously enabled stoichiometric purification of WDR5-interacting proteins<sup>28</sup> also specifically retrieved *HOTTIP* RNA (Supplementary Fig. 9). Knockdown of WDR5 broadly inhibited expression of 5' *HOXA* genes, and also abrogated *HOTTIP* transcription, demonstrating mutual interdependence between *HOTTIP* RNA and WDR5 (Fig. 4d).

*HOTTIP* RNA seems to regulate genes *in cis*, due to its low copy number, distance dependence of *HOXA* target gene activation on endogenous *HOTTIP*, and the physical proximity of *HOTTIP* and its target genes as seen in 5C. Indeed, ectopic expression of *HOTTIP* RNA by retroviral transduction of lung fibroblasts, which do not express *HOTTIP*, failed to activate expression of distal *HOXA* genes, and did not change H3K4me3 and H3K27me3 patterns across *HOXA* (Supplementary Fig. 10). Moreover, in foreskin fibroblasts that express endogenous *HOTTIP*, ectopic *HOTTIP* expression did not induce 5' *HOXA* genes, nor rescue the effects of depleting endogenous nascent *HOTTIP* RNA (Supplementary Fig. 11). The lack of response in foreskin fibroblasts is notable because endogenous *HOTTIP* RNA is active in these cells, indicating that the protein partners of *HOTTIP* are all present and target genes are receptive. Ectopically expressed *HOTTIP* RNA, being transcribed from retroviral insertion sites scattered randomly in the genome, may not be able to find 5' *HOXA* genes. In contrast, endogenous *HOTTIP* RNA is directly positioned near the 5' *HOXA* genes by chromosomal looping, allowing interaction and control.



**Figure 4 | *HOTTIP* RNA programs active chromatin via WDR5.**

**a**, Summary of RNA-protein interaction studies. Each of the indicated recombinant protein was purified and used to retrieve purified *HOTTIP* RNA or control histone RNA *in vitro*. Only GST-WDR5 specifically retrieved *HOTTIP*. **b**, *HOTTIP* RNA binds directly and specifically to WDR5. Left, purified GST and GST-WDR5 are visualized by SDS-PAGE and Coomassie Blue staining. Right, retrieved RNAs are quantified by qRT-PCR. **c**, *HOTTIP* RNA binds specifically to WDR5 in cells. Immunoprecipitation (IP) of endogenous WDR5 protein from PC3 (prostate) and T24 (bladder) carcinoma cells specifically retrieved *HOTTIP*, but not control IPs with IgG or chromatin binder SIRT6. U1 spliceosomal RNA served as negative control. **d**, WDR5 is required for 5' *HOXA* gene expression, including *HOTTIP* RNA. **e**, *HOTTIP* RNA recruitment potentiates transcription. Left, the *BoxB* tethering system. *BoxB*-RNA specifically binds  $\lambda$ N fused to GAL4 DNA binding domain, recruiting the complex to a UAS-luciferase reporter gene. After transient transfection, IP of GAL4- $\lambda$ N specifically retrieves *BoxB*-*HOTTIP* RNA. Right, luciferase activity after co-transfection of the indicated constructs. \* $P < 0.05$  Student's *t*-test comparing *BoxB*-LacZ versus *BoxB*-*HOTTIP*. Sloped triangle indicates increasing input of plasmids encoding ncRNAs. Means  $\pm$  s.d. ( $n = 3$ ) are shown for all panels.

To test the requirement of an exogenous targeting mechanism, we engineered an allele of *HOTTIP* RNA that can be artificially recruited to a reporter gene. Addition of five copies of the *BoxB* RNA element<sup>29</sup> to *HOTTIP* RNA allows the fusion transcript to be recruited to the  $\lambda$ N RNA binding domain fused to a GAL4 DNA-binding domain (Fig. 4e). Recruitment of *HOTTIP* RNA to a silent GAL4 promoter is not sufficient to initiate transcription, but can significantly boost transcription if the promoter is also bound by WDR5 and transcriptionally active (Fig. 4e). By uncoupling the sites of *HOTTIP* transcription versus *HOTTIP* RNA function, this experiment indicates that the proximity of *HOTTIP* RNA—rather than the act of transcription—maintains target gene expression. To further support the functionality of *HOTTIP* RNA, deletion analysis identified a  $\sim 1$  kb domain in the 5' of *HOTTIP* RNA (*HOTTIP*<sup>Exons 1–2</sup>) that retains WDR5 binding activity (Supplementary Fig. 12a). Enforced overexpression of *HOTTIP*<sup>Exons 1–2</sup>



in foreskin fibroblasts inhibited 5' *HOXA* gene expression in an apparently dominant negative manner (Supplementary Fig. 12b).

In summary, *HOTTIP* RNA is a key locus control element of *HOXA* genes and distal identity. Chromosomal looping brings *HOTTIP* RNA in close proximity to the 5' *HOXA* genes. *HOTTIP* transcription acts as a switch to produce *HOTTIP* lincRNA, which binds to and targets WDR5–MLL complexes to the 5' *HOXA* locus, yielding a broad domain of H3K4me3 and transcription activation (Supplementary Fig. 13). The mutual interdependence between *HOTTIP* RNA and WDR5 creates a positive feedback loop that maintains the ON state of the locus. These findings provide an integrated view linking three dimensional genome organization to dynamic programming of chromatin states, and ultimately to developmental pattern formation.

H3K4 methylation is a feature of almost all transcribed genes, and MLL family proteins are involved in many cell fate decisions in development and disease<sup>24</sup>. Our findings suggest that additional lincRNAs, especially those associated with enhancers or enhancer-like activities<sup>8–10</sup>, may also be involved in gene activation by programming active chromatin states, and highlight WDR5 and other WD40 repeat proteins as candidate adaptors that link chromatin remodelling complexes to lincRNAs. *Cis*-restricted lincRNAs may be ideally suited to link chromosome structure and gene expression. Because such lincRNA can only act on its neighbours in space, information in higher order chromosomal looping can be faithfully transmitted to chromatin modification via RNA recruitment of enzymatic activities, and thus into gene expression.

## METHODS SUMMARY

High throughput chromosome conformation capture (5C) was performed on foreskin and lung fibroblasts, as well as foreskin fibroblasts treated with control or siRNA against *HOTTIP* RNA, as described<sup>13</sup>. siRNA knockdown experiments on cultured human fibroblasts and qPCR were performed as described previously<sup>12</sup>. ChIP-chip was performed as described<sup>12</sup> using ultra-high-density *HOX* tiling arrays. Full-length *HOTTIP* RNA was cloned by 5' and 3' rapid amplification of cloned/cDNA ends (RACE). Single-molecule RNA-fluorescent in situ hybridization (FISH) was performed using a pool of fluorescently-labelled oligonucleotides specific to *HOTTIP* RNA. *In vivo* *HOTTIP* RNA knockdown in chick was accomplished by microinjecting retroviruses carrying shRNA into prospective wing and leg buds and the animals were harvested at 2, 4 and 9 weeks post injection for RNA *in situ* hybridization, immunohistochemistry and whole-mount limb analysis, respectively. RNA-immunoprecipitation with WDR5 was performed as described<sup>12</sup>. Tethering experiments were done in 293T cells with co-transfections of various constructs containing an upstream activating sequence (UAS)-luciferase reporter, *GAL4*-WDR5, *BoxB* alone, and *BoxB* fused to full-length *HOTTIP* or full-length *LacZ*; cells were lysed 48 h after transfection and luciferase activity was determined.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 February 2010; accepted 12 January 2011.

Published online 20 March 2011.

- Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10**, 155–159 (2009).
- Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136**, 629–641 (2009).
- Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to *Ultrabithorax*. *Science* **311**, 1118–1123 (2006).
- Petruk, S. *et al.* Transcription of *bxd* noncoding RNAs promoted by trithorax represses *Ubx* in *cis* by transcriptional interference. *Cell* **127**, 1209–1221 (2006).
- Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18**, 1433–1445 (2008).
- Dean, A. On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* **22**, 38–45 (2006).
- Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human  $\beta$ -globin locus. *Genes Dev.* **11**, 2494–2509 (1997).

- De Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46–58 (2010).
- Chang, H. Y. Anatomic demarcation of cells: genes to patterns. *Science* **326**, 1206–1207 (2009).
- Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
- Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128**, 735–745 (2007).
- Zhang, X. *et al.* A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human *HOXA* cluster. *Blood* **113**, 2526–2534 (2009).
- Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
- Kmita, M., Fraudeau, N., Herault, Y. & Duboule, D. Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. *Nature* **420**, 145–150 (2002).
- Harpavat, S. & Cepko, C. L. RCAS-RNAi: a loss-of-function method for the developing chick retina. *BMC Dev. Biol.* **6**, 2 (2006).
- Nelson, C. E. *et al.* Analysis of *Hox* gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
- Kmita, M. *et al.* Early developmental arrest of mammalian limbs lacking *HoxA/HoxD* gene function. *Nature* **435**, 1113–1116 (2005).
- Small, K. M. & Potter, S. S. Homeotic transformations and limb defects in *HoxA11* mutant mice. *Genes Dev.* **7**, 2318–2328 (1993).
- Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capecci, M. R. Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375**, 791–795 (1995).
- Fromental-Ramain, C. *et al.* *Hoxa-13* and *Hoxd-13* play a crucial role in the patterning of the limb autopod. *Development* **122**, 2997–3011 (1996).
- Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25**, 15–30 (2007).
- Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29**, 6074–6085 (2009).
- Guenther, M. G. *et al.* Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl Acad. Sci. USA* **102**, 8603–8608 (2005).
- Krajewski, W. A., Nakamura, T., Mazo, A. & Canaani, E. A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell. Biol.* **25**, 1891–1899 (2005).
- Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121**, 859–872 (2005).
- Baron-Benhamou, J., Gehring, N. H., Kulozik, A. E. & Hentze, M. W. Using the  $\lambda$ N peptide to tether proteins to RNAs. *Methods Mol. Biol.* **257**, 135–154 (2004).
- Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nature Methods* **6**, 690–691 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Tabin for chick *Hox* gene probes, M. Scott and members of our labs for input, and M. Lin for use of the confocal microscope and imaging expertise. Supported by grants from the California Institute for Regenerative Medicine (H.Y.C., J.W.), the National Institutes of Health (HG003143 to J.D.), and the Scleroderma Research Foundation (H.Y.C.). K.C.W. is a recipient of a Dermatology Foundation Career Development Award. J.D. is a recipient of the W. M. Keck Foundation Distinguished Young Scholar Award. H.Y.C. and M.L. are Early Career Scientists of the Howard Hughes Medical Institute.

**Author Contributions** K.C.W., R.A.G. and H.Y.C. initiated the project; K.C.W. and H.Y.C. designed the experiments; K.C.W., Y.W.Y., B.L., A.S., R.C.-Z., B.R.L., A.P., R.A.F., J.D. and J.A.H. conducted the experiments and analysed the data; Y.C. and M.L. purified the recombinant proteins; J.W. provided antibodies and cell lines; K.C.W. and H.Y.C. prepared the manuscript with inputs from all co-authors.

**Author Information** Sequence for human *HOTTIP* RNA has been deposited with GenBank under the accession number GU724873. Microarray data are deposited in Gene Expression Omnibus (GEO) under accession number GSE26540. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.Y.C. ([howchang@stanford.edu](mailto:howchang@stanford.edu)).

## METHODS

**Cells.** Primary human fibroblasts derived from different anatomic sites were as described<sup>12,31–36</sup>. Primary human fibroblasts in culture retain their positional identity and have been used to examine chromatin states associated with positional memory, which have been confirmed *in vivo*<sup>33,34,37</sup>.

**Chromatin immunoprecipitation followed by microarray analysis.** ChIP-chip was performed using anti-H3K27me3 (Abcam), anti-H3K4me3 (Abcam), anti-H3K4me2 (Abcam), anti-histone H3 (Abcam), anti-PolII (Abcam), anti-MLL1 (gift of R. Roeder), and anti-WDR5<sup>28</sup> antibodies as previously described<sup>12</sup>. Chromatin from each indicated cell type or RNAi treatment is split into multiple tubes and subject to ChIP with different antibodies in parallel. Retrieved DNA and input chromatin were competitively hybridized to custom tiling arrays interrogating human *HOX* loci at 5-bp resolution as previously described<sup>12</sup>.

**5C analysis of the ENM010 HoxA1 region.** 5C primers were designed at HindIII restriction sites using 5C primer design tools previously developed<sup>13</sup> and made available online at <http://my5C.umassmed.edu> (ref. 30). Reverse primers were designed for fragments overlapping a known transcription start site from GENCODE transcripts<sup>38</sup>, or overlapping a start site as experimentally determined by CAGE Tag data of the ENCODE pilot project<sup>39</sup>. Forward primers were designed for all other HindIII restriction fragments. Primers were excluded if highly repetitive sequences prevented the design of a sufficiently unique 5C primer. Primer settings were: U-BLAST: 3; S-BLAST: 130; 15-MER: 1320; MIN\_FSIZE: 40; MAX\_FSIZE: 50000; OPT\_TM: 65; OPT\_PSIZE: 40. DNA sequence of the universal tails of forward primers was CCTCTCTATGGGCAGTCGGTGAT; DNA sequence for the universal tails of reverse primers was AGAGAATGAGGAACC CGGGGCAG. A 6-base barcode was included between the specific part of the primers and the universal tail. In total 17 reverse primers and 90 forward primers were designed in the 500 kb *HoxA1* locus (ENM010) and hence a total of 1,530 *cis* interaction were interrogated in this region. Primer sequences are available separately (Supplementary Table 1).

3C was performed with HindIII as previously described<sup>40</sup> separately for fetal lung and foreskin fibroblasts (FB) and also for the control and *HOTTIP* knock-down foreskin FBs. For the 5C reaction, a total of 107 forward and reverse primers of *HoxA1* region were mixed with either the ENCODE random region (ENr) primer pool comprising of 2,673 forward and 523 reverse primers (covering 30 additional ENCODE regions) or the ENr313 primer pool comprising of 57 forward and 58 reverse primers (covering 1 additional ENCODE region). 5C was then performed in 10 reactions each containing an amount of 3C library that represents 200,000 genome equivalents and 1 fmol of each primer. The 5C analysis of *HoxA1* region was carried out in two biological replicates of fetal lung and foreskin FBs. 5C ligation products were amplified using a pair of universal primers that recognize the common tails of the 5C forward and reverse primers described above and pooled together. To facilitate paired-end DNA sequence analysis on the Illumina GA2 platform, paired-end adaptor oligonucleotides were ligated to the 5C library using the Illumina PE protocol and PCR amplification of the library was carried out for 18 cycles with Illumina PCR primer PE 1.0 and 2.0. The 5C library was then sequenced on the Illumina GA2 platform generating 36 base paired end reads. For fetal lung FBs we obtained 7,625,276 and 10,947,424 mapped reads for two biological replicates of which 1,339,861 and 242,301 could be specifically mapped back to interactions within ENM010 using Novoalign (<http://www.novocraft.com>), respectively. For two biological replicates of foreskin FBs we obtained 7,311,386 and 5,731,107 mapped reads of which 2,752,789 and 66,769 could be mapped back to the ENM010 region, respectively. In the case of the knockdown study, control green fluorescent protein (GFP) knockdown foreskin FB 5C library yielded 4,909,482 mapped reads whereas *HOTTIP* knockdown foreskin FB had 5,565,389 mapped reads of which 39,168 and 38,950 could be mapped back to ENM010 for control GFP and *HOTTIP* knockdown, respectively. In the set with fetal lung and foreskin fibroblast samples, 5C for ENM010 was multiplexed for deep sequencing with 5C of one other region, ENr313; in the set containing the knockdown samples, ENM010 was multiplexed with 5C of 30 other genomic regions. The different extent of multiplexing resulted in different number of sequencing reads mapping back to ENM010. In all instances the mappable reads were proportional to the degree of multiplexing, indicating equivalent library quality despite different read numbers. Supplementary Table 2 outlines the library composition of each experiment. The heat maps are scaled as follows—for Fig. 1a, distal (foreskin) FBs: 262–17,467, proximal (lung) FBs: 7–5,846; for Supplementary Fig. 6, siGFP: 1–100, siHOTTIP: 1–100. Raw data from the 5C experiments used to generate the binned heat maps in Fig. 1a and Supplementary Fig. 6 can be found in Supplementary File 1. Raw data are available by request.

**HOTTIP cloning, sequence and expression analysis.** We previously identified a portion of *HOTTIP* as a non protein-coding transcribed region named nCHOXA13-96 (ref. 12). This region also overlaps expressed sequence tag (EST) clone AK093987 that was previously observed to be expressed in cancer cell lines

derived from posterior anatomic sites<sup>41</sup>. 5' and 3' RACE (RLM Race kit, Applied Biosystems/Ambion) showed full-length *HOTTIP* RNA to be 3,764 nucleotides, extending the known transcribed region by more than 1,400 bases. BLAST and BLAT confirmed that portions of *HOTTIP* are well conserved in mammals and even in avians but had no protein coding potential. Full-length *HOTTIP* RNA sequence has been deposited at NCBI (accession number GU724873). qRT-PCR with SYBR Green was conducted as recommended by the manufacturer (Agilent Technologies). Primer sequences specific for *HOTTIP* were CCTAAAGCCACGC TTCTTTG (*HOTTIP*-F) and TGCAGGCTGGAGATCCTACT (*HOTTIP*-R). For Supplementary Fig. 11, endogenous nascent *HOTTIP* was distinguished from ectopic *HOTTIP* expressed from cDNA using primers that spanned intron–exon junctions.

**Strand-specific RT-PCR.** RNA extracted from primary foreskin fibroblasts was reverse transcribed (SuperScript III, Invitrogen) using combinations of the previously described *HOTTIP*-specific primers *HOTTIP*-F and/or *HOTTIP*-R as diagrammed in Supplementary Fig. 1b. Resulting cDNA was then PCR-amplified using both *HOTTIP*-F and *HOTTIP*-R primers to visually determine strand specificity.

**HOTTIP transcript count per cell.** The level of *HOTTIP* transcript per cell was calculated from the level of *HOTTIP* in 500,000 cells. Full-length *HOTTIP* in pcDNA3.1+ was assayed by qPCR using primers *HOTTIP*-F and *HOTTIP*-R at predetermined concentrations in triplicate to generate a linear amplification curve dependent on the moles of template DNA (Supplementary Fig. 2). The qRT-PCR value from 500,000 foreskin fibroblasts was determined and plotted, and the corresponding total molecules of transcript was divided by 500,000 to determine the approximate number of transcripts per cell.

**Single-molecule RNA fluorescence *in situ* hybridization (RNA-FISH).** Single molecule RNA-FISH was performed as described in ref. 42 with the following modifications: the amount of hybridization solution per chamber was doubled to allow for proper coating of the chamber and the amount of glucose-oxidase buffer was tripled to assist in image acquisition. Images were acquired using an Olympus FV1000 confocal microscope within 2 h of the addition of the glucose-oxidase buffer.

**RNA interference.** Primary foreskin fibroblasts were transfected with siRNAs targeting *HOTTIP* and *WDR5* using Lipofectamine 2000 (Invitrogen) as per manufacturer's instructions. Total RNA was harvested 48–72 h later using TRIzol (Invitrogen) and RNeasy Mini Kits (Qiagen) as previously described<sup>34</sup>. For the intronic *HOTTIP* knockdown experiment in Supplementary Figure 11, a pool of 10 siRNAs (Supplementary Table 3) targeting intronic regions in *HOTTIP* were transfected into foreskin fibroblasts, and RNA isolated as above.

**Generation of shRNAs against chicken *HOTTIP*.** A reporter construct encoding a GFP–chicken *HOTTIP* fusion transcript was used in a small-scale screen to identify highly effective shRNA constructs. Eleven shRNAs targeting conserved regions of chicken *HOTTIP* were designed and inserted into the pSMP system (Thermo/Open Biosystems). The reporter construct and shRNA constructs were cotransfected into Phoenix cells, and *HOTTIP* transcript levels were analysed via reduced GFP fluorescence and by qRT-PCR. Three shRNAs that were effective *in vitro* were then cloned into RCAS vector for studies in chick embryos<sup>18</sup>.

**Chick RNAi.** RCAS *HOTTIP* hairpin and RCAS AP viruses were made by transfecting DF-1 cells with viral DNA. Transfected DF-1 cells were grown and passaged, after which the virus-containing supernatant was collected, concentrated and titred. Fertilized chicken eggs were incubated in a humidified rotating incubator at 37 °C until they reached Hamilton/Hamburger stage 10. Eggs were then windowed to expose the embryos. After gently removing the vitelline membrane, chicken embryos were microinjected with RCAS-*HOTTIP* hairpins and RCAS-AP viruses at the prospective wing and leg buds. All viral stocks have titres of  $1 \times 10^8$  IU ml<sup>-1</sup>, and each limb was injected five times. The infected embryos were allowed to incubate at 37 °C and were harvested 2 or 4 days after injection to detect viral infection by immunohistochemistry. Total RNA was extracted from injected forelimbs, and RT-PCR analysis was performed 4 days after injection. Chicken embryos were harvested 9 days post-injection to carry out whole-mount Alcian blue staining. A total of 50 animals were injected.

Hairpin sequences for chick *HOTTIP* were TGCTGTTGACAGTGAGCGAC CCGAAGATGTGTCTGATTGTAGTGAAGCCACAGATGTACAAATCAGACACATCTTCGGCTGCCTACTGCTCGGA (2-2-1), TGCTGTTGACAGTGAGCGCGCTCTGCTCTCTCTCTCTAGTGAAGCCACAGATGTAGAGAGAGAGGAGAGCAGAGCGATGCCTACTGCCTCGGA (3-2-1), and TGCTGTTGACAGTGAGCGAATCCTTAATCGAATCTGATTTAGTGAAGCCACAGATGTAAATCAGATTCGATTAAGGATCTGCCTACTGCCTCGGA (4-4-1).

**HOTTIP overexpression.** Full-length *HOTTIP* and a truncated transcript consisting of exons 1 and 2 (*HOTTIP*<sup>Exons 1–2</sup>) were cloned into the LZRS vector (gift of P. Khavari), and then transfected into Phoenix cells (gift of G. Nolan) to generate amphotropic retroviruses. Primary human fibroblasts were infected with either LZRS-full length *HOTTIP* (lung), LZRS-truncated *HOTTIP* (foreskin), or LZRS-GFP (both lung and foreskin), then passaged over 60 days, with periodic

testing of *HOXA* and *HOTTIP* expression by qRT-PCR. These cells were used in the rescue experiments depicted in Supplementary Fig. 11.

**GST pull-down.** Full-length *HOTTIP*, truncated *HOTTIP* containing exons 1 and 2 (*HOTTIP*<sup>Exons 1–2</sup>), and histone H2B1 mRNA were transcribed *in vitro* using T7 polymerase according to manufacturer's instructions (Promega), denatured, and refolded in folding buffer (100 mM KCl, 10 mM MgCl<sub>2</sub>, Tris pH 7.0). GST-tagged WDR5, C-terminal MLL1, RBBP5/Ash2L and TRF1 were expressed in *Escherichia coli* and purified as described<sup>43</sup>. Each GST-fusion protein was bound to glutathione beads (Amersham/GE Healthcare) and blocked with excess yeast total mRNA in PB100 buffer (20 mM HEPES pH 7.6, 100 mM KCl, 0.05% NP40, 1 mM DTT, 0.5 mM PMSF) for 1 h at room temperature. Beads were then incubated with either *in-vitro*-transcribed *HOTTIP* or histone H2B1 mRNA for 45 min at room temperature. After three washes in PB200 buffer (20 mM HEPES pH 7.6, 200 mM KCl, 0.05% NP40, 1 mM DTT, 0.5 mM PMSF), bound RNAs were extracted and analysed by qRT-PCR, as previously described.

**RNA immunoprecipitation.** HeLa-WDR5-Flag cells: 48 h after Lipofectamine 2000-mediated transfection of *HOTTIP* into HeLa WDR5-Flag cells (approximately 10<sup>7</sup> cells), total protein was extracted as previously described, with modifications<sup>44</sup>. Briefly, cells were resuspended in Buffer A (10 mM HEPES pH 7.5, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, 0.5 mM DTT, 1.0 mM PMSF), lysed in 0.25% NP40, and fractionated by low speed centrifugation. The nuclear fraction was resuspended and lysed in Buffer C (20 mM HEPES pH 7.5, 10% glycerol, 0.42 M KCl, 4 mM MgCl<sub>2</sub>, 0.5 mM DTT, 1.0 mM PMSF). Combined nuclear and cytoplasmic fractions were immunoprecipitated with mouse anti-Flag M2 monoclonal antibody (Sigma) or mouse IgG affixed to agarose beads (Sigma) for 3 to 4 h at 4 °C. Beads were washed four times with wash buffer (50 mM TrisCl pH 7.9, 10% glycerol, 100 mM KCl, 5 mM MgCl<sub>2</sub>, 10 mM β-mercaptoethanol, 0.1% NP40). After elution using Flag peptide (Sigma), co-immunoprecipitated RNA was extracted and analysed by qRT-PCR.

Endogenous WDR5 and SIRT6 RIP: cellular fractions were isolated as above and incubated with the anti-WDR5 (ref. 31) or anti-Sirt6 (ab62739, Abcam) antibodies overnight at 4 °C. Samples were washed in wash buffer, and co-immunoprecipitated RNA was extracted and analysed by qRT-PCR.

**RNA chromatography.** Full-length *in-vitro*-transcribed *HOTTIP* RNA was conjugated to adipic acid dehydrazide agarose beads as described<sup>45</sup>. The complexed beads were incubated with whole cell lysates from HeLa WDR5-Flag cells, washed, and bound proteins visualized by western blotting.

**BoxB tethering assay.** 293T cells were grown to about 50% confluence in 6-well plates on the day of transfection. Using Lipofectamine 2000 (Invitrogen), a plasmid

encoding a luciferase gene under the control of five tandem GAL4 UAS sites were co-transfected with plasmids encoding GAL4-WDR5, GAL4-λN (the 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N) peptide fused to a C-terminal GFP tag, *BoxB* (containing five repeats of the λN-specific 19 nucleotide binding site), *BoxB* fused to full-length *LacZ*, or *BoxB* fused to full-length *HOTTIP*. Cells were lysed 48 h after transfection, and luciferase assay kit (Promega) was used to determine relative levels of the luciferase gene product, following the manufacturer's protocol.

31. Chang, H. Y. *et al.* Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl Acad. Sci. USA* **99**, 12877–12882 (2002).
32. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2**, 206–214 (2004).
33. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
34. Rinn, J. L., Bonde, C., Gladstone, H. B., Brown, P. O. & Chang, H. Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2**, e119 (2006).
35. Rinn, J. L. *et al.* A dermal *HOX* transcriptional program regulates site-specific epidermal fate. *Genes Dev.* **22**, 303–307 (2008).
36. Rinn, J. L. *et al.* A systems biology approach to anatomic diversity of skin. *J. Invest. Dermatol.* **128**, 776–782 (2008).
37. Soshnikova, N. & Duboule, D. Epigenetic temporal control of mouse *Hox* genes in vivo. *Science* **324**, 1320–1323 (2009).
38. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (Suppl 1), S4 (2006).
39. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
40. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols* **2**, 988–1002 (2007).
41. Sasaki, Y. T., Sano, M., Kin, T., Asai, K. & Hirose, T. Coordinated expression of ncRNAs and *HOX* mRNAs in the human *HOXA* locus. *Biochem. Biophys. Res. Commun.* **357**, 724–730 (2007).
42. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5**, 877–879 (2008).
43. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31–40 (1988).
44. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).
45. Michlewski, G. & Caceres, J. F. RNase-assisted RNA chromatography. *RNA* **16**, 1673–1678 (2010).



# Molecular regulation of sexual preference revealed by genetic studies of 5-HT in the brains of male mice

Yan Liu<sup>1,2\*</sup>, Yun'ai Jiang<sup>1,3\*</sup>, Yunxia Si<sup>1</sup>, Ji-Young Kim<sup>4</sup>, Zhou-Feng Chen<sup>4</sup> & Yi Rao<sup>1,5</sup>

Although the question of to whom a male directs his mating attempts<sup>1,2</sup> is a critical one in social interactions, little is known about the molecular and cellular mechanisms controlling mammalian sexual preference. Here we report that the neurotransmitter 5-hydroxytryptamine (5-HT) is required for male sexual preference. Wild-type male mice preferred females over males, but males lacking central serotonergic neurons lost sexual preference although they were not generally defective in olfaction or in pheromone sensing. A role for 5-HT was demonstrated by the phenotype of mice lacking tryptophan hydroxylase 2 (Tph2), which is required for the first step of 5-HT synthesis in the brain. Thirty-five minutes after the injection of the intermediate 5-hydroxytryptophan (5-HTP), which circumvented Tph2 to restore 5-HT to the wild-type level, adult *Tph2* knockout mice also preferred females over males. These results indicate that 5-HT and serotonergic neurons in the adult brain regulate mammalian sexual preference.

Interactions between members of the opposite sex are essential for sexually reproducing animals. Evolutionary benefits have been proposed for homo- and bisexual traits<sup>1,2</sup>, which exist in many animals<sup>2</sup> from American bulls<sup>3</sup> to Japanese rhesus monkeys<sup>4</sup>. Studies of animals with different sexual preferences are essential for understanding the seemingly simple decision of a male to court a female.

Research in *Drosophila* has uncovered genes required for *Drosophila* courtship preference, but none of their homologues have been shown to affect mammalian sexual preference. Research in mammals has demonstrated that pheromone sensing in the periphery is important for sexual preference. Male mice lacking *Trpc2* (*Trpc2*<sup>-/-</sup>), which encodes a channel expressed in the vomeronasal organ, mounted other males, emitted ultrasonic vocalizations (USVs) towards males and were less aggressive towards males<sup>5,6</sup>. However, understanding of the central mechanisms for sexual preference remains limited.

The neurotransmitter 5-HT has been implicated in male sexual behaviours such as erection, ejaculation and orgasm in mice and humans<sup>7,8</sup>. Depletion of 5-HT by treating animals with p-chlorophenylalanine (pCPA) or tryptophan-free diets induced male–male mounting<sup>9–11</sup>. However, pCPA treatment was thought to increase sexual activity whereas its effect on sexual preference has not been investigated. Interpretation of pCPA results was complicated further by the lack of specificity: pCPA may affect noradrenaline and dopamine at higher concentrations<sup>12</sup>.

Almost all serotonergic neurons in the brain were missing from embryogenesis to adulthood in *Lmx1b* conditional knockout mice in which the floxed *Lmx1b* allele was deleted by *ePet1*-Cre<sup>13</sup>. We compared the behaviours of male mice of different genotypes: *ePet1*-Cre/*Lmx1b*<sup>fllox</sup>/*Lmx1b*<sup>fllox</sup> as homozygous mutants (*Lmx1b*<sup>-/-</sup>); their littermates *ePet1*-Cre/*Lmx1b*<sup>fllox/+</sup> as heterozygous mutants (*Lmx1b*<sup>+/-</sup>); and *Lmx1b*<sup>fllox</sup>/*Lmx1b*<sup>fllox</sup> without *ePet1*-Cre as the wild type (*Lmx1b*<sup>+/+</sup>). We also used *ePet1*-Cre without *Lmx1b*<sup>fllox</sup> as a control.

We tested first how a male responded in his home cage when a wild-type target C57 male was introduced. Compared to the *ePet1*-Cre,

*Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> controls, *Lmx1b*<sup>-/-</sup> mice showed significantly more mounting of male intruders (Fig. 1 and Supplementary Movie 1; see Supplementary Data 1 for numbers of mice used and statistics for all figures). The percentage of males who mounted target males was significantly higher in *Lmx1b*<sup>-/-</sup> males than *ePet1*-Cre, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>+/+</sup> males (Fig. 1a). *Lmx1b*<sup>-/-</sup> males mounted with a shorter latency (Fig. 1b), higher frequency (Fig. 1c) and longer duration (Fig. 1d). These results show that the absence of serotonergic neurons in the brain increased male–male mounting.

A sexually dimorphic behavioural response of males is to emit 30–110 kHz USVs when they encounter female mice or pheromones, which may function as love songs to facilitate female receptivity<sup>14</sup>. *Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>-/-</sup> males were similar in USV emission towards females (Fig. 1e–g). However, the percentage of *Lmx1b*<sup>-/-</sup> males emitting USV towards males was significantly higher than that of *ePet1*-Cre, *Lmx1b*<sup>+/+</sup> or *Lmx1b*<sup>+/-</sup> males (Fig. 1f). Numbers of USV ‘syllables’ emitted towards females were similar among *ePet1*-Cre, *Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>-/-</sup> males (Fig. 1g). *Lmx1b*<sup>-/-</sup> males emitted more USV ‘syllables’ towards males than *ePet1*-Cre, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup>. The number of USV emissions by *Lmx1b*<sup>-/-</sup> males towards males was approximately 720 times higher than that of *Lmx1b*<sup>+/+</sup> males (Fig. 1g).

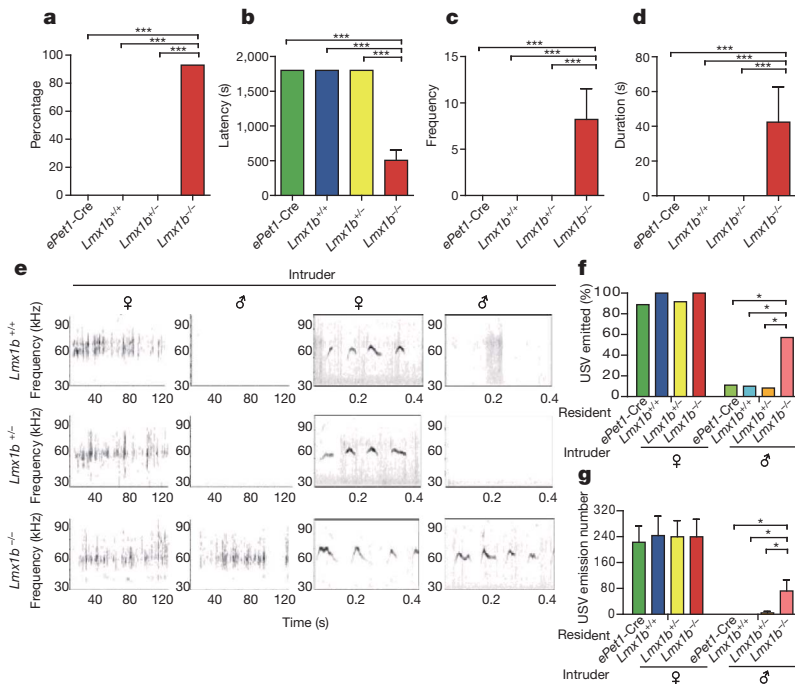
Although *Lmx1b*<sup>-/-</sup> males still emitted more USVs towards females, the preference for females over males was significantly reduced: the ratio of USVs towards females over that for males was only 3 for *Lmx1b*<sup>-/-</sup> males, significantly reduced from 1,002 for *ePet1*-Cre males, 2,438 for *Lmx1b*<sup>+/+</sup> males and 52 for *Lmx1b*<sup>+/-</sup>.

In the mating choice assay, an oestrous female C57 target mouse and a sexually naive male C57 target mouse were introduced into the home cage of a test male. Wild-type males preferred to mount female targets (Fig. 2a): a higher percentage of *Lmx1b*<sup>+/+</sup> (or *ePet1*-Cre, *Lmx1b*<sup>+/-</sup>) males mounted female targets than male targets (Supplementary Movie 2). However, the percentage of *Lmx1b*<sup>-/-</sup> males mounting females was not significantly different from that mounting males. *ePet1*-Cre, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males mounted female targets with a shorter latency, higher frequency and longer duration than male targets (Fig. 2b, d, e), whereas *Lmx1b*<sup>-/-</sup> males mounted males and females with similar latencies, frequencies and durations (Supplementary Movies 2 and 3). Thus, elimination of serotonergic neurons led to a loss of sexual preference in mounting.

Further analyses were carried out to detect a change in sexual preference separate from an increase in sexual drive: (1) in the mating choice assay, all *ePet1*-Cre, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males mounted females before males, whereas 46.2% of *Lmx1b*<sup>-/-</sup> mounted males first (Fig. 2c); (2) the mounting frequency ratio of *Lmx1b*<sup>-/-</sup> males in the mating choice assay (female mounting frequency – male mounting frequency)/(female + male mounting) (that is, (♀ – ♂)/♂ + ♀) was significantly different from *ePet1*-Cre, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males (Fig. 2f); and (3) when a test male was presented only with an oestrous female target, *Lmx1b*<sup>-/-</sup> males were not statistically significant

<sup>1</sup>National Institute of Biological Sciences, Beijing 102206, China. <sup>2</sup>Graduate School of the Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China. <sup>3</sup>Institute of Neuroscience, Shanghai Institute of Biological Sciences, and Graduate School of the China Academy of Science, China. <sup>4</sup>Departments of Anesthesiology, Psychiatry and Developmental Biology, and the Pain Center, Washington University, School of Medicine, St Louis, Missouri 63110, USA. <sup>5</sup>Peking University School of Life Sciences, State Key Laboratory of Membrane Biology, Beijing 100871, China.

\*These authors contributed equally to this work.



**Figure 1 | Male-male mounting and USV by mice lacking central serotonergic neurons.** **a–g**, Numbers of mice used and statistical analysis are all included in Supplementary Data 1. **a–d**, A test male was presented in its home cage with an adult wild-type male and its behaviour was recorded for 30 min (all data shown as mean  $\pm$  s.e.m.). Compared with *Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> or *ePet1-Cre, Lmx1b*<sup>-/-</sup> males mounted males at a higher percentage (**a**), lower latency (**b**), higher frequency (**c**) and for a longer duration (**d**). **e**, Typical USV patterns emitted by males when presented with female or male intruders. The two left panels show USVs in 2 min, whereas the two right

panels show parts of USV graphs at higher magnifications. **f**, Female intruders elicited USV from almost all *ePet1-Cre, Lmx1b*<sup>-/-</sup>, *Lmx1b*<sup>+/+</sup> or *Lmx1b*<sup>+/-</sup> males. Male intruders elicited USVs more from *Lmx1b*<sup>-/-</sup> males than from *ePet1-Cre, Lmx1b*<sup>+/+</sup> or *Lmx1b*<sup>+/-</sup> males. **g**, The number of USVs emitted by *Lmx1b*<sup>-/-</sup> males towards males is higher than those by *ePet1-Cre, Lmx1b*<sup>+/+</sup> or *Lmx1b*<sup>+/-</sup> males, whereas *ePet1-Cre, Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>-/-</sup> males were similar in USVs towards females. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

different from wild-type and heterozygous males in male–female mounting (Supplementary Fig. 1).

We tested male mice for their preference of pheromones present in the genitals or the bedding. In the genital odour preference assay<sup>15</sup>, a slide with one half smeared with female genitals and the other half with male genitals was presented to a test male. The total time spent sniffing both halves of the slide was reduced in *Lmx1b*<sup>-/-</sup> males (Supplementary Fig. 2a). *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> littermates spent significantly more time sniffing female than male genital odour, whereas *Lmx1b*<sup>-/-</sup> males spent equal time sniffing female and male genital odours (Fig. 3a). *Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>-/-</sup> were similar in the amount of time spent sniffing male genital odour. Female genital odour sniffing time was less in *Lmx1b*<sup>-/-</sup> males than in *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> littermates (Fig. 3a). The genital odour preference ratio ( $Q - \sigma / Q + \sigma$ ) of *Lmx1b*<sup>-/-</sup> males was significantly lower than those of *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males (Fig. 3b). Compared with *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males, a significantly higher percentage (62.5%) of *Lmx1b*<sup>-/-</sup> males spent more time sniffing male than female genital odour (Fig. 3c).

In the bedding preference assay<sup>16</sup>, the total time spent over male and female bedding was similar among *ePet1-Cre, Lmx1b*<sup>+/+</sup>, *Lmx1b*<sup>+/-</sup> and *Lmx1b*<sup>-/-</sup> males (Supplementary Fig. 2b). *ePet1-Cre, Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males spent significantly more time above female than male bedding whereas *Lmx1b*<sup>-/-</sup> males spent equal time above female and male beddings (Fig. 3d). Compared with *ePet1-Cre, Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males, *Lmx1b*<sup>-/-</sup> males spent more time above male bedding and less time above female bedding. The bedding preference ratio of *Lmx1b*<sup>-/-</sup> males was significantly lower than those of *ePet1-Cre, Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>+/-</sup> males (Fig. 3e). The percentage of males who spent more time above male bedding was significantly higher in *Lmx1b*<sup>-/-</sup> males (58.8%) than those in *ePet1-Cre* (0%), *Lmx1b*<sup>+/+</sup> (6.3%) or *Lmx1b*<sup>+/-</sup> (12.5%) males (Fig. 3f).

Thus, in both the genital odour and bedding assays, *Lmx1b*<sup>-/-</sup> males had lost preference for female pheromones over male pheromones: in the genital odour preference assay, *Lmx1b*<sup>-/-</sup> males showed decreased sniffing time for female genital odour; in the bedding preference assay, *Lmx1b*<sup>-/-</sup> males showed increased time spent over male bedding and decreased time over female bedding.

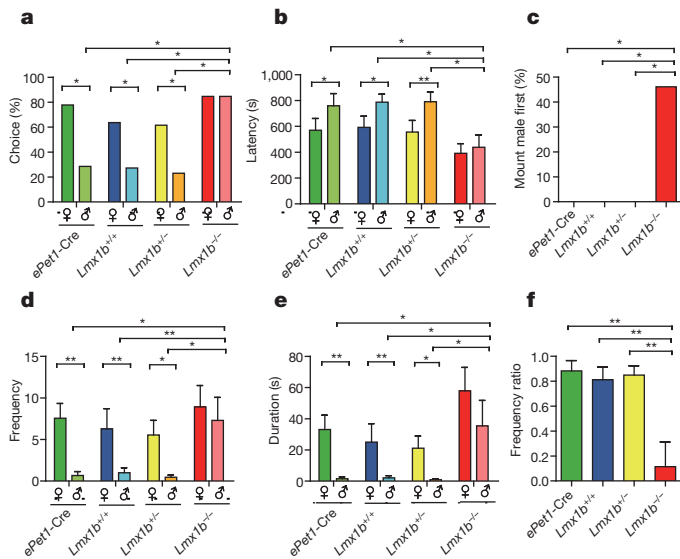
Multiple assays involving odour or pheromone sensing were carried out to test for possible changes in olfaction. In the sesame oil preference assay<sup>17</sup>, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>-/-</sup> males were indistinguishable in spending significantly more time with sesame than air (Supplementary Fig. 3a). In the fox urine avoidance assay<sup>18</sup>, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>-/-</sup> males were also similar (Supplementary Fig. 3b). Thus, *Lmx1b*<sup>-/-</sup> males were not defective in either innate attractive or avoidance response.

In the social approach assay<sup>19</sup>, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>-/-</sup> males were similar in spending more time close to a strange male than the empty chamber (Supplementary Fig. 3c).

In the social recognition assay<sup>20</sup>, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>-/-</sup> males spent a similar amount of time exploring the first intruder at initial presentation, displayed social habituation towards the familiar intruder over the next three presentations and displayed dishabituation when a new intruder was introduced (Fig. 4a).

An operant conditioning assay was used to test whether *Lmx1b*<sup>-/-</sup> males could distinguish between male and female pheromones<sup>21</sup>. Two arms of a T maze were supplied with the odour of either female or male urine. Electroshock was applied in such a way that the test mice had to run or stay in the same arm depending on the urine. Over 3 days of training, *Lmx1b*<sup>+/+</sup> and *Lmx1b*<sup>-/-</sup> males were similar in learning to avoid punishment (Fig. 4b). Thus, no olfactory defects for general odours or pheromones were detected in *Lmx1b*<sup>-/-</sup> males.

Results from *Lmx1b*<sup>-/-</sup> mice indicate a role for serotonergic neurons. To study the role of 5-HT, we used mice unable to synthesize



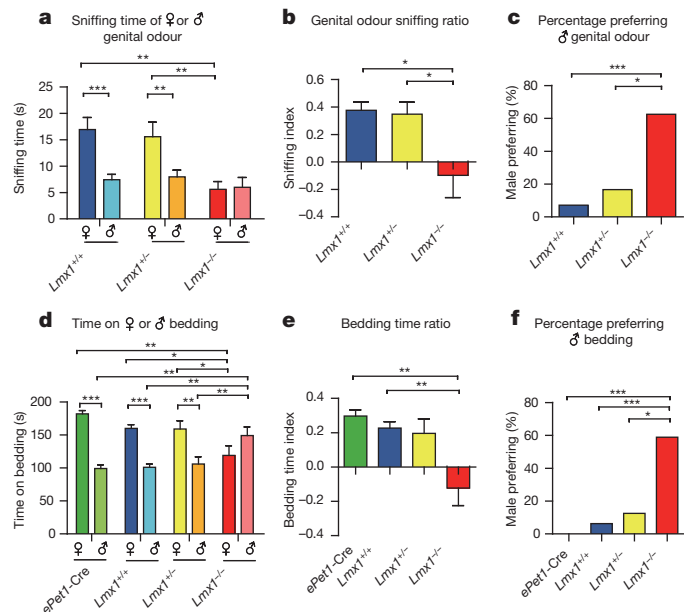
**Figure 2 | Lack of sexual preference by mice without central serotonergic neurons.** **a–f**, Each test male was presented with a male and an oestrous female, and its mating choice was analysed for 15 min. **a**, More *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males mounted female than male targets. A similar percentage of *Lmx1b<sup>-/-</sup>* males mounted females and males. **b**, *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males mounted female targets faster than male targets. Mounting latencies of *Lmx1b<sup>-/-</sup>* males for females and males were similar. **c**, More than 40% of *Lmx1b<sup>-/-</sup>* males but none of the *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* or *Lmx1b<sup>+/-</sup>* males chose a male as their first mounting target. **d**, *ePet1-Cre* males mounted females significantly more often than males, as did *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males. *Lmx1b<sup>-/-</sup>* males mounted females as often as males ( $P > 0.05$ , *t*-test). **e**, *ePet1-Cre* males spent more time mounting females than males, as did *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males. *Lmx1b<sup>-/-</sup>* males did not show differences in mounting males or females. **f**, The mounting frequency ratio of *Lmx1b<sup>-/-</sup>* was different from that of *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>*. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

5-HT in the brain. 5-HT is synthesized in two steps: tryptophan is converted by a Tph into 5-HTP, which is converted into 5-HT by 5-hydroxytryptophan decarboxylase and aromatic L-amino-acid decarboxylase.

There are two Tph enzymes: Tph2 is required centrally and Tph1 peripherally. We have generated *Tph2<sup>-/-</sup>* mice (J.-Y.K. *et al.*, manuscript in preparation), which were viable<sup>22–24</sup>. High-performance liquid chromatography (HPLC) analysis showed that the 5-HT level was significantly reduced in the brains of *Tph2<sup>-/-</sup>* males (Supplementary Fig. 4a). Male–male mounting (Supplementary Movie 4) was significantly higher in *Tph2<sup>-/-</sup>* males than either *Tph2<sup>+/+</sup>* or heterozygous *Tph2<sup>+/-</sup>* males: the percentage was significantly higher, duration longer, latency shorter and frequency higher (Supplementary Fig. 4b, c and Fig. 5a, b). In the bedding preference assay, both *Tph2<sup>+/+</sup>* and *Tph2<sup>+/-</sup>* males preferred female over male bedding, whereas *Tph2<sup>-/-</sup>* males showed no preference (Fig. 5c). In the genital odour preference assay, both *Tph2<sup>+/+</sup>* and *Tph2<sup>+/-</sup>* males preferred female over male genital odour, but *Tph2<sup>-/-</sup>* males showed no preference (Fig. 5d).

When presented with an oestrous female target, male–female mounting was not significantly changed in *Tph2<sup>-/-</sup>* males (Supplementary Fig. 5). In mating choice, *Tph2<sup>-/-</sup>* males had lost preference for females over males in percentage, latency, frequency and duration (Supplementary Fig. 6a, b, d, e). No control males mounted target males before females, whereas more than 40% of *Tph2<sup>-/-</sup>* males mounted males first (Supplementary Fig. 6c). The mounting frequency ratio of *Tph2<sup>-/-</sup>* males was significantly different from those of *Tph2<sup>+/+</sup>* and *Tph2<sup>+/-</sup>* males (Supplementary Fig. 6f).

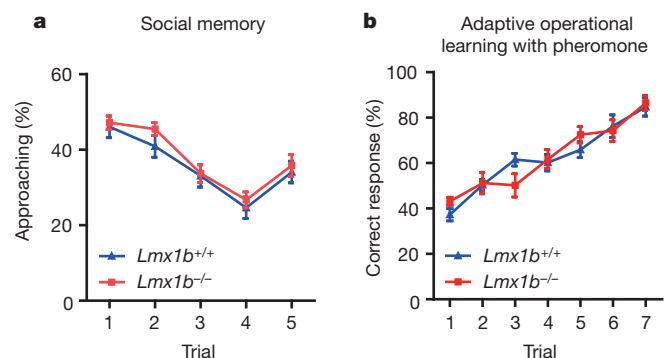
*Lmx1b<sup>-/-</sup>* and *Tph2<sup>-/-</sup>* mice lack 5-HT from embryogenesis. To study the role of 5-HT in adulthood, we took two complementary



**Figure 3 | Loss of sexual preference for genital odour and bedding by males without central serotonergic neurons.** **a**, *Lmx1b<sup>+/+</sup>* males spent more time sniffing female than male genital odour, as did *Lmx1b<sup>+/-</sup>* males. *Lmx1b<sup>-/-</sup>* males spent a similar amount of time on female and male genital odour. Three groups were not significantly different in male genital odour sniffing time but *Lmx1b<sup>-/-</sup>* males spent less time sniffing female genital odour than the other two groups. **b**, Sniffing ratio of *Lmx1b<sup>-/-</sup>* males was significantly different from *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males ( $P < 0.05$  for *Lmx1b<sup>+/+</sup>* versus *Lmx1b<sup>-/-</sup>*,  $P < 0.05$  for *Lmx1b<sup>+/-</sup>* versus *Lmx1b<sup>-/-</sup>*,  $P > 0.05$  for *Lmx1b<sup>+/+</sup>* versus *Lmx1b<sup>+/-</sup>*; one-way ANOVA). **c**, Compared with *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>*, a higher percentage of *Lmx1b<sup>-/-</sup>* males spent more time sniffing male than female genital odour. **d**, *ePet1-Cre* males spent more time above female bedding than male bedding, as did *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>* males. *Lmx1b<sup>-/-</sup>* males spent a similar amount of time above female and male bedding. Compared with *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>*, *Lmx1b<sup>-/-</sup>* males spent less time above female bedding but more time above male bedding. **e**, The bedding time ratio of *Lmx1b<sup>-/-</sup>* was different from *ePet1-Cre* and *Lmx1b<sup>+/+</sup>*. **f**, Compared with *ePet1-Cre*, *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>+/-</sup>*, a significantly higher percentage of *Lmx1b<sup>-/-</sup>* males spent more time above male bedding. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

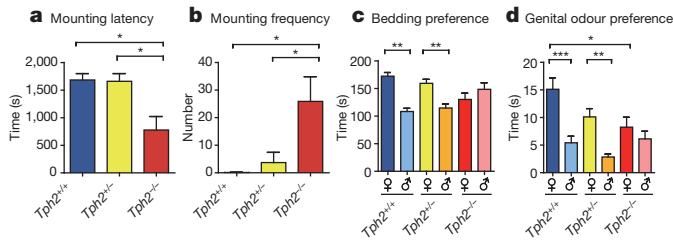
approaches: first, we depleted 5-HT from adult mice pharmacologically with pCPA<sup>25</sup>; then we attempted to rescue the phenotype of adult *Tph2<sup>-/-</sup>* mutants.

Adult C57BL/6J males were injected with either pCPA or saline for three consecutive days. 5-HT level was significantly reduced by pCPA



**Figure 4 | Odour discrimination.** **a**, Both *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>-/-</sup>* males showed habituation and dishabituation in sniffing time. No statistical difference was found between *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>-/-</sup>* males at any point. **b**, After seven training sessions with male and female urine, no significant difference was detected between *Lmx1b<sup>+/+</sup>* and *Lmx1b<sup>-/-</sup>* males at any point.





**Figure 5 | Brain chemistry and behaviours of *Tph2* knockout males.**

**a, b,** Compared with *Tph2*<sup>+/+</sup> and *Tph2*<sup>+/-</sup>, *Tph2*<sup>-/-</sup> males showed a shorter latency (**a**) and higher frequency in mounting males (**b**). **c,** Both *Tph2*<sup>+/+</sup> and *Tph2*<sup>+/-</sup> males significantly preferred female over male bedding, whereas *Tph2*<sup>-/-</sup> males did not show a preference between male and female bedding. **d,** Both *Tph2*<sup>+/+</sup> and *Tph2*<sup>+/-</sup> males significantly preferred female over male genital odour, whereas *Tph2*<sup>-/-</sup> males did not show a preference between male and female genital odour. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001.

(Supplementary Fig. 7). pCPA-treated males showed shorter latency, higher frequency and longer duration than control males in mounting target males (Supplementary Fig. 8a–d), and lost bedding preference (Supplementary Fig. 8e, f).

To test whether 5-HTP injection into adult mice could rescue the *Tph2*<sup>-/-</sup> phenotype, we examined first whether 5-HTP could rescue 5-HT synthesis in *Tph2*<sup>-/-</sup> males and found that 5-HT levels were restored 35 min after intraperitoneal injection of 5-HTP but not saline (Fig. 6a and Supplementary 9a, b).

5-HTP significantly reduced male–male mounting of *Tph2*<sup>-/-</sup> males: the percentage was decreased, latency increased, frequency decreased and duration shortened; all returning to wild-type levels (Fig. 6b, c and Supplementary Fig. 9c, d). 5-HTP rescued the loss of sexual preference in mounting latency, frequency and duration in the mating choice assay (Supplementary Fig. 10a–c) and the bedding preference of *Tph2*<sup>-/-</sup> males (Fig. 6d and Supplementary Fig. 9e).

When a test male was presented with a target female, *Tph2*<sup>-/-</sup> males were similar to wild-type and heterozygous males in mounting percentage, latency, frequency and duration (Supplementary Figs 5, 11). 5-HTP injection into *Tph2*<sup>-/-</sup> males did not affect male–female mounting (Supplementary Fig. 11), although 5-HTP injection into wild-type males reduced male–female mounting. Because 5-HTP injection in wild-type males increased the level of 5-HT beyond the wild-type level (Supplementary Fig. 9a, b), it indicated a dosage-sensitive effect of 5-HT: 5-HT at concentrations above the wild-type level inhibited male–female mounting, but 5-HT concentrations between the wild-type and *Tph2*<sup>-/-</sup> levels did not affect male–female mounting.

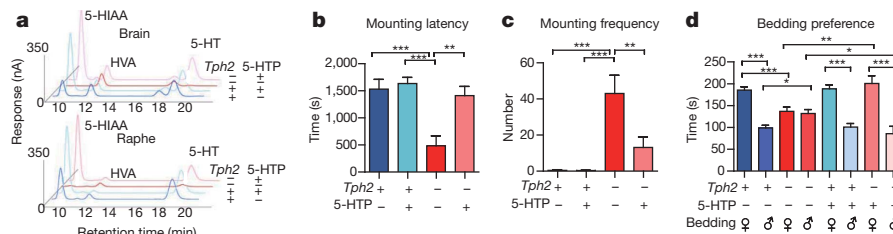
We conclude that central serotonergic signalling is crucial for male sexual preference in mice. This is the first time, to our knowledge, that a neurotransmitter in the brain has been demonstrated to be important in mammalian sexual preference. Previous studies in mammals have implicated 5-HT and dopamine in male sexual behaviours, but neither has been demonstrated to have any role in sexual preference: dopamine is thought to facilitate male sexual behaviours whereas 5-HT is

thought to inhibit sexual behaviours<sup>7–11,26</sup>. Our studies have established a role for 5-HT in male sexual preference. Multiple results showed a loss in sexual preference beyond or separate from hypersexuality: (1) the ratio of male–male and male–female interactions was repeatedly measured to analyse sexual preference (Figs 2f, 3b, e, 5c, d, 6d and Supplementary Figs 6f, 8f, 9e, 10d); (2) *Lmx1b*<sup>-/-</sup> males showed increased USVs towards males but not towards females (Fig. 1g); (3) in mating choice, the latency, frequency and duration of *Lmx1b*<sup>-/-</sup> males to mount males, but not to mount females, was changed (Fig. 2a, b, d, e); (4) in bedding preference, *Lmx1b*<sup>-/-</sup> (Fig. 3d) and *Tph2*<sup>-/-</sup> males (Figs 5c, 6d) showed an increase in time spent over male bedding but a decrease in time over female bedding; (5) wild-type males always mounted females before males but a significant fraction of *Lmx1b*<sup>-/-</sup> or *Tph2*<sup>-/-</sup> males mounted males first (Fig. 2c and Supplementary Fig. 6c); (6) in the genital odour preference assay, both *Lmx1b*<sup>-/-</sup> (Fig. 3a) and *Tph2*<sup>-/-</sup> (Supplementary Fig. 5d) males showed a decrease in time on female genital odour, which could not be explained by hypersexuality; and (7) when presented with an oestrous target female, neither *Lmx1b*<sup>-/-</sup> males (Supplementary Fig. 1) nor *Tph2*<sup>-/-</sup> males (Supplementary Fig. 5) were different from wild-type males.

Increased sexual drive was observed in males lacking 5-HT when they were tested in the presence of live target males and females (Supplementary Fig. 6). This has been noted before in mice defective for *Trpc2* and vomeronasal organ olfaction<sup>5,6</sup>. *Trpc2*<sup>-/-</sup> males have been previously reported to have lost male–female preference in mating choice<sup>5,6</sup>. *Trpc2*<sup>-/-</sup> males showed increased mounting of both males and females (figure 2c in ref. 6). The conclusion of a loss in sexual preference in *Trpc2*<sup>-/-</sup> males was inferred from a relative change: *Trpc2*<sup>-/-</sup> males showed a 2-fold preference for females over males whereas the wild-type showed a 10-fold preference. The phenotypes reported here for *Lmx1b*<sup>-/-</sup>, *Tph2*<sup>-/-</sup> males and pCPA-treated males were stronger than for *Trpc2*<sup>-/-</sup> males in mating choice: these males did not show significant preference for females (Fig. 2 and Supplementary Fig. 6).

At present, it is not known whether 5-HT regulates the vomeronasal organ pathway in pheromone sensing or acts further downstream in behavioural decisions. Differences have been noted between *Trpc2* and *Lmx1b* in the brain: aggression was largely lost in *Trpc2*<sup>-/-</sup>, but not *Lmx1b*<sup>-/-</sup>, mice (data not shown). It is more likely that 5-HT regulates central decision-making than influencing peripheral olfaction. However, we cannot completely rule out the possibility that 5-HT regulates a specific innate olfactory pathway processing sexual information<sup>27</sup>. In mice, it will be interesting to identify specific subsets of serotonergic neurons and serotonergic receptors involved in sexual preference.

An unavoidable question raised by our findings is whether 5-HT has a role in sexual preference in other animals. In a positron emission tomography study of humans, the response of heterosexual men to the selective serotonin reuptake inhibitor (SSRI) fluoxetine was found to be different from that of homosexual men<sup>28</sup>. SSRIs inhibited compulsive sexual behaviours in homosexual and bisexual men<sup>29</sup>. However, so far, none of these studies has investigated whether 5-HT has a role in



**Figure 6 | 5-HTP rescue of chemical and behavioural deficits in *Tph2* knockout mice.** **a,** Levels of 5-HT and 5-hydroxyindoleacetic acid (5-HIAA) were analysed in *Tph2*<sup>+/+</sup> and *Tph2*<sup>-/-</sup> males 35 min after injection of either 5-HTP (40 mg kg<sup>-1</sup> body weight) or control saline. **b, c,** Male–male mounting

in *Tph2*<sup>-/-</sup> mice was significantly rescued by 5-HTP: the latency was lengthened and frequency reduced. **d,** Bedding preference was monitored between 35 and 40 min after injection. 5-HTP could significantly restore the preference of female over male bedding by *Tph2*<sup>-/-</sup> males.

sexual preference. Attempts have been made to map genetic loci affecting human sexuality<sup>30</sup>, although specific genes have not been identified. Our discovery of a role for serotonergic signalling in mouse sexual preference should stimulate further studies into the role of 5-HT in sexual interactions in particular and roles of neurotransmitters in mammalian social relationships in general.

## METHODS SUMMARY

We used conditional knockout mice for *Lmx1b* and knockout mice for *Tph2*. Levels of 5-HT in these mice and their heterozygous and wild-type littermates were measured by HPLC. Most of the behavioural assays were similar to established methods.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 12 August 2010; accepted 14 January 2011.

Published online 23 March 2011.

- Trivers, R. L. Parent-offspring conflict. *Am. Zool.* **14**, 249–264 (1974).
- Sommer, V. & Vasey, P. L. *Homosexual Behaviour in Animals: An Evolutionary Perspective* (Cambridge Univ. Press, 2006).
- Price, E. O. & Wallach, S. J. Development of sexual and aggressive behaviors in Hereford bulls. *J. Anim. Sci.* **69**, 1019–1027 (1991).
- Erwin, J. & Maple, T. Ambisexual behavior with male–male anal penetration in male rhesus monkeys. *Arch. Sex. Behav.* **5**, 9–14 (1976).
- Stowers, L., Holy, T. E., Meister, M., Dulac, C. & Koentges, G. Loss of sex discrimination and male–male aggression in mice deficient for TRP2. *Science* **295**, 1493–1500 (2002).
- Leybold, B. G. *et al.* Altered sexual and social behaviors in *trp2* mutant mice. *Proc. Natl Acad. Sci. USA* **99**, 6376–6381 (2002).
- Hull, E. M., Muschamp, J. W. & Sato, S. Dopamine and serotonin: influences on male sexual behavior. *Physiol. Behav.* **83**, 291–307 (2004).
- Hull, E. M. & Dominguez, J. M. Sexual behavior in male rodents. *Horm. Behav.* **52**, 45–55 (2007).
- Ferguson, J. *et al.* “Hypersexuality” and behavioral changes in cats caused by administration of p-chlorophenylalanine. *Science* **168**, 499–501 (1970).
- Malmnäs, C. & Meyerson, B. p-Chlorophenylalanine and copulatory behaviour in the male rat. *Nature* **232**, 398–400 (1971).
- Salis, P. & Dewsbury, D. p-Chlorophenylalanine facilitates copulatory behaviour in male rats. *Nature* **232**, 400–401 (1971).
- Dailly, E., Chenu, F., Petit-Demouliere, B. & Bourin, M. Specificity and efficacy of noradrenaline, serotonin depletion in discrete brain areas of Swiss mice by neurotoxins. *J. Neurosci. Methods* **150**, 111–115 (2006).
- Zhao, Z.-Q. *et al.* *Lmx1b* is required for maintenance of central serotonergic neurons and mice lacking central serotonergic system exhibit normal locomotor activity. *J. Neurosci.* **26**, 12781–12788 (2006).
- Guo, Z. & Holy, T. E. Sex selectivity of mouse ultrasonic songs. *Chem. Senses* **32**, 463–473 (2007).
- Ferkin, M. H. & Li, H. Z. A battery of olfactory-based screens for phenotyping the social and sexual behaviors of mice. *Physiol. Behav.* **85**, 489–499 (2005).
- Moncho-Bogani, J., Lanuza, E., Hernandez, A., Novejarque, A. & Martez-Garc, F. Attractive properties of sexual pheromones in mice: innate or learned? *Physiol. Behav.* **77**, 167–176 (2002).
- Burwash, M. D., Tobin, M. E., Woolhouse, A. D. & Sullivan, T. P. Laboratory evaluation of predator odors for eliciting an avoidance response in roof rats (*Rattus rattus*). *J. Chem. Ecol.* **24**, 49–66 (1998).
- Blanchard, D. *et al.* Failure to produce conditioning with low-dose trimethylthiazole or cat feces as unconditioned stimuli. *Behav. Neurosci.* **117**, 360–368 (2003).
- Nadler, J. J. *et al.* Automated apparatus for quantitation of social approach behaviors in mice. *Genes Brain Behav.* **3**, 303–314 (2004).
- Ferguson, J. N. *et al.* Social amnesia in mice lacking the oxytocin gene. *Nature Genet.* **25**, 284–288 (2000).
- Yan, Z. *et al.* Precise circuitry links bilaterally symmetric olfactory maps. *Neuron* **58**, 613–624 (2008).
- Gutknecht, L. *et al.* Deficiency of brain 5-HT synthesis but serotonergic neuron formation in *Tph2* knockout mice. *J. Neural Transm.* **115**, 1127–1132 (2008).
- Savelieva, K. V. *et al.* Genetic disruption of both tryptophan hydroxylase genes dramatically reduces serotonin and affects behavior in models sensitive to antidepressants. *PLoS ONE* **3**, e3301 (2008).
- Alenina, N. *et al.* Growth retardation and altered autonomic control in mice lacking brain serotonin. *Proc. Natl Acad. Sci. USA* **106**, 10332–10337 (2009).
- Koe, B. K. & Weissman, A. p-Chlorophenylalanine: a specific depletor of brain serotonin. *J. Pharmacol. Exp. Ther.* **154**, 499–516 (1966).
- Gawienowski, A. M. & Hodgen, G. D. Homosexual activity in male rats after p-chlorophenylalanine: effects of hypophysectomy and testosterone. *Physiol. Behav.* **7**, 551–555 (1971).
- Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
- Kinnunen, L., Moltz, H., Metz, J. & Cooper, M. Differential brain activation in exclusively homosexual and heterosexual men produced by the selective serotonin reuptake inhibitor, fluoxetine. *Brain Res.* **1024**, 251–254 (2004).
- Wainberg, M. *et al.* A double-blind study of citalopram versus placebo in the treatment of compulsive sexual behaviors in gay and bisexual men. *J. Clin. Psychiatry* **67**, 1968–1973 (2006).
- Mustanski, B. S. *et al.* A genomewide scan of male sexual orientation. *Hum. Genet.* **116**, 272–278 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to E. S. Deneris for *ePet1-Cre* mice; to R. Johnson for *Lmx1b<sup>fl</sup>* mice; to M. Luo for discussions; to Z. Yan and Y. Lu for the operant conditioning apparatus; to X. Wang and Y. Wan for help with HPLC; to J. Lang and J. Yin for mouse breeding and genotyping; to P. Ding, P. Wang, H. Lu and X. Wang for technical assistance; to L. Zhao, Z. Qiu and H. Jing for animal caring; and to the Ministry of Science and Technology (973 program 2010CB833901) and Beijing Municipal Commission on Science and Technology for grant support (to Y.R.), and the NIH for grant support (to Z.-F.C.).

**Author Contributions** Y.R. conceived the project, Y.R., Y.L. and Y.J. designed the experiments, Y.L., Y.J. and Y.S. performed the experiments, J.-Y.K. and Z.-F.C. contributed the *Tph2* knockout mutants, Y.R., Y.L. and Y.J. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to Y.R. ([yrao@pku.edu.cn](mailto:yrao@pku.edu.cn)).

## METHODS

**Mouse stocks.** *ePet1*-Cre mice were a gift from E. S. Deneris and the floxed *Lmx1b* mice were a gift from R. Johnson. *Tph2* knockout mice were generated by deleting exon 5, which encodes the tryptophan hydroxylase domain (for details see J.-Y.K. *et al.*, manuscript submitted). Mice were weaned at the age of 21 days. Mice were maintained on a 12 h light, 12 h dark schedule and housed initially in groups of five up to the tenth week and then singly housed until the end of experiments. Food and water were provided *ad libitum*. Room temperature was  $23 \pm 1^\circ\text{C}$ . Humidity was 40–60%. All test mice were 12–16 weeks old. The target mice were 11–13 weeks old.

**Mouse genotyping.** Genomic DNA was extracted from mouse tail tissues at the day of weaning. Mutant mice were generated by crossing *ePet1*-Cre mice with floxed *Lmx1b* mice and following intercross within the F1 generation mice. Littermates used in the tests were of the same sex and similar body weight as the knockout mice. The primers were: AGGCTCCATCCATCTTCTC (floxed *Lmx1b1*); CCACAATAAGCAAGAGGCAC (floxed *Lmx1b2*); ATTTGCCTGCA TTACCGTCTG (Cre1); CAGCATTTGCTGCTCACTTGGTC (Cre2).

Immunocytochemical analysis with anti-5-HT antibodies confirmed that 5-HT-positive neurons were absent in *Lmx1b* knockout mice (data not shown).

The *Tph2* line was maintained by crossing heterozygotes. Littermates included wild-type, heterozygotes and homozygous knockout mice. The primers for genotyping were: GGGCATCTCAGACGTAGTAG; GGGCTGCCGATAGTAA CAC; GCAGCCAGTAGACGTCTCTTAC.

**Measurement of 5-HT.** The levels of 5-HT and its metabolites were separated by HPLC and measured by an electrochemical detector in samples from adult male mice. In 5-HTP rescue experiments, mice were injected with  $40\text{ mg kg}^{-1}$  5-HTP or saline (both at the volume of  $5\text{ ml kg}^{-1}$ ). They were euthanized 35 min later. The brain was dissected and the raphe region was isolated on ice. Samples were weighed before ultrasonication. Monoamines were extracted by perchloric acid. The sample was filtrated by  $0.22\text{ }\mu\text{m}$  filter before being injected into RP-HPLC (ESA). Noradrenaline, 3,4-dihydroxyphenylacetic acid (DOPAC), dopamine, HIAA, homovanillic acid (HVA) and 5-HT were measured by an electrochemical detector. Their concentrations were calculated by CoulArray software (ESA) based on standard samples. Values of amine per wet tissue weight are shown in the final figures.

**Order of behavioural assays.** Male mutant mice and their littermates at 12–13 weeks of age and of similar body weight were sexually naive and group-housed with same-sex mice before 10 weeks of age. After 2 weeks of single housing, mice were tested in the following order: bedding preference, male–male resident–intruder assay, mating choice assay, sexual behaviours with an oestrous female, bedding preference again (no difference was observed with results from the first bedding preference). Mice were given one week of rest between each test. For *Lmx1b* mice, the same group of mice were used in male–male mounting, mating choice and male–female mounting. For *Tph2* mice, a different group were used for male–female mounting. Sexually experienced mice were used for USV, social approach, habituation and olfactory learning assays. Sexually naive mice were used for urine preference and olfactory tests.

**Resident–intruder tests.** All test mice were sexually naive. The bedding of the test mice had not been changed for at least 4 days. Intruder mice were 11–13 weeks old, sexually naive and group-housed C57Bl/6J males. All activities within a test were recorded by an infrared camera (Sony Video Recorder, DCR-HC26C). Mounting latency, mounting frequency and total duration of mounting within 30 min were measured.

**Mating choice assay.** Beddings of test mice had not been changed for at least four days. A group-housed sexually naive 11–13 week-old C57Bl/6J male and a sexually naive oestrous 10-week-old female C57Bl/6J female were introduced into the cage of each test male. Each assay lasted 15 min after the target mice were introduced. All activities were recorded by an infrared camera. The latency, frequency and duration of mounting of male or female targets were analysed.

**Sexual behaviours with females.** An oestrous female was presented to a test male and video was recorded for 30 min using an infrared camera. The latency, frequency and duration of male mounting of the female were analysed.

**USVs.** Tests were carried out with singly housed adult males during the dark phase in the home cage. UltraSoundGate 116–200 system (Avisoft) was used to record the ultrasound. We recorded the background sound for 1 min before a stimulus mouse of 10–13 weeks old was introduced. The recording lasted for 2 min. Recorded data was analysed with SASLab (Avisoft)<sup>5</sup>. Sounds over the frequency range of 30–110 kHz were analysed. Profiles of background noise created by mouse movement were very different from USVs. To confirm that the resident mouse was the source of USVs, we recorded from assays in which either the resident or the intruder mouse was devocalized. We were able to record robust USVs (presented in our figures) only when the intruder mouse was devocalized and not when the resident mouse was devocalized.

**Genital odour preference assay.** This assay was modified from a previously described procedure<sup>15</sup>. The anogenital area scent from a male was rubbed on

the left or right side of a clean glass microscope slide while the anogenital area scent from a female was rubbed on the other side of the slide. Five seconds later, the slide was hung in the middle of the cage by a clamp. The slides were  $\sim 5\text{ cm}$  over the bedding. Activities of the test mice were recorded for 3 min by an infrared camera and the sniff time on the scent portion on either side was analysed as was the amount of time a test male licked the slide or its nose touched the slide.

**Bedding preference assay.** Bedding from group-housed adult C57Bl/6J males or females was not changed for 4 days. Ten grams of male or female bedding were put in one side on the bottom of a cage in an area of  $11.5 \times 17\text{ cm}^2$ . Male and female beddings were prevented from mixing by a plastic bar of 6 cm. The size of cage was  $29 \times 17 \times 15\text{ cm}$  (length  $\times$  width  $\times$  height)<sup>16</sup>. A grid of plastic bars separated the test mice from the bedding on the bottom of the cage. The bars were 5 mm wide with 5 mm intervals. The test mouse was put into the cage to be familiarized with the cage without bedding for 5 min before the mice were taken out and the bedding and a clean grid was put into the cage. After each assay, the cage was washed with water and then alcohol to remove odour.

**Olfactory learning assay.** We employed a T maze in which electric shock could be applied to either side of the horizontal chamber as described previously<sup>21</sup>. Briefly, there was a door at the intersection of the horizontal and vertical chambers. The horizontal chamber of  $8 \times 8 \times 60\text{ cm}^3$  was divided into three parts: a left arm of  $8 \times 8 \times 23\text{ cm}$ , a right arm of  $8 \times 8 \times 23\text{ cm}$  and a middle zone of  $8 \times 8 \times 14\text{ cm}$ . Each test mouse was introduced into the vertical chamber of the T maze. After it entered the horizontal chamber, the door between the vertical and horizontal chambers was closed and the mouse was allowed to walk within the horizontal chamber. The mouse was not allowed to stay in the middle zone for longer than 8 s, otherwise it would be punished with electroshock. The position of the test mouse was monitored by a video recorder. Urine samples were collected from more than 20 C57Bl/6J males or females and stored at  $-20^\circ\text{C}$ . A 1.5 ml urine sample was used for each test. The odour of male or female urine was puffed into the left or right arm of the horizontal chamber and expired from the middle zone. Odour was presented for 50 s. We trained the test male mouse with electroshock to stay in the arm with female odour and to avoid the arm with male odour. The mouse had to make a decision to stay in or leave the arm when an odour was presented. Each training session of 18 trials lasted for 30 min. Every mouse was given 6 training sessions over 3 days before the final test. There were 10 trials in the final test. The percentages of correct choices in every training session and the final test were analysed.

**Innate behavioural responses to odours.** The set-up is the same as that for the olfactory learning assay, except that no electroshock was applied. Sexually naive males (mutants or littermates) of 10–16 weeks old were tested for their choices of fox urine versus air, or sesame oil versus air. Fox urine was used to test the innate avoidance of a predator's odour. Fox urine was diluted at two concentrations ( $60\times$  and  $20\times$ ). The main air flow velocity was  $250\text{ l h}^{-1}$ . The air flow through fox urine was  $70\text{ ml min}^{-1}$  or  $210\text{ ml min}^{-1}$ , respectively. The time that mice spent in the empty arm or the fox urine arm was recorded by Matlab software. Sesame oil diluted  $83\times$  was used to test innate attraction to food. Time spent in the air arm or the sesame oil arm was recorded by Matlab software.

**Social approach.** The social approach experiment was tested in a modified T-shaped box. There was a small cage separated by wire at each end of the arms in the horizontal chamber. A test mouse was allowed to habituate for five minutes before an unfamiliar target male was randomly placed in one of the small cages. The target mouse could be seen, smelled and heard, but could not be touched. The test mouse was allowed to move in the box for 5 min. Its location was video recorded and analysed by a computer.

**Social memory.** Singly housed adult males were tested in the dark phase and in the room where they were reared. Ovariectomized C57Bl/6J females were used as stimulus mice<sup>20</sup>. They were ovariectomized at 6 weeks old and used 2 weeks later. A stimulus mouse was introduced into the cage housing a test mouse for 1 min and then was removed. After an interval of 10 min, the same stimulus female was introduced again for 1 min. The stimulus mouse was presented four times. On the fifth time, a new stimulus mouse was introduced for 1 min. The behaviour of test mice was videotaped and time spent on body sniffing was analysed.

**5-HT depletion by pCPA treatment.** Male C57Bl/6J mice of 11–13 weeks of age were used. They were injected with either  $500\text{ mg kg}^{-1}$  of pCPA (Sigma, C6506) or saline control for 3 consecutive days after 4 days of being singly housed. Animals were tested with adult C57 female mice. Mice that did not show mounting behaviour in 15 min were discarded. Mice that qualified were then singly housed for 1 week before social behaviour testing and their bedding was not changed. Animals were randomly divided into pCPA or saline treatment groups. pCPA was suspended in 1% Tween saline at a concentration of  $50\text{ mg ml}^{-1}$ . The pCPA group were injected intraperitoneally with pCPA ( $10\text{ ml kg}^{-1}$ ) at 72, 48 and 24 h before testing. The control group received 1% Tween saline. Resident–intruder and mating choice assays were carried out. Behavioural tests were performed in the dark.



# Local, persistent activation of Rho GTPases during plasticity of single dendritic spines

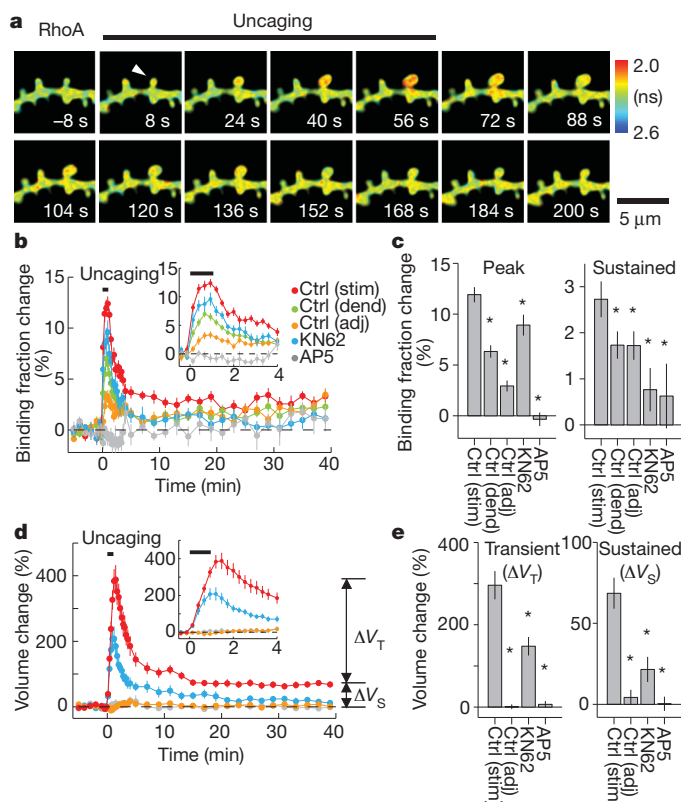
Hideji Murakoshi<sup>1</sup>, Hong Wang<sup>1</sup> & Ryohei Yasuda<sup>1,2</sup>

The Rho family of GTPases have important roles in the morphogenesis of the dendritic spines<sup>1–3</sup> of neurons in the brain and synaptic plasticity<sup>4–9</sup> by modulating the organization of the actin cytoskeleton<sup>10</sup>. Here we used two-photon fluorescence lifetime imaging microscopy<sup>11–13</sup> to monitor the activity of two Rho GTPases—RhoA and Cdc42—in single dendritic spines undergoing structural plasticity associated with long-term potentiation in CA1 pyramidal neurons in cultured slices of rat hippocampus. When long-term volume increase was induced in a single spine using two-photon glutamate uncaging<sup>14,15</sup>, RhoA and Cdc42 were rapidly activated in the stimulated spine. These activities decayed over about five minutes, and were then followed by a phase of persistent activation lasting more than half an hour. Although active RhoA and Cdc42 were similarly mobile, their activity patterns were different. RhoA activation diffused out of the stimulated spine and spread over about 5  $\mu\text{m}$  along the dendrite. In contrast, Cdc42 activation was restricted to the stimulated spine, and exhibited a steep gradient at the spine necks. Inhibition of the Rho–Rock pathway preferentially inhibited the initial spine growth, whereas the inhibition of the Cdc42–Pak pathway blocked the maintenance of sustained structural plasticity. RhoA and Cdc42 activation depended on  $\text{Ca}^{2+}$ /calmodulin-dependent kinase (CaMKII). Thus, RhoA and Cdc42 relay transient CaMKII activation<sup>13</sup> to synapse-specific, long-term signalling required for spine structural plasticity.

Previous studies using two-photon fluorescence lifetime imaging microscopy (2pFLIM) and two-photon glutamate uncaging revealed the spatiotemporal dynamics of the signalling proteins CaMKII and HRas (also known as transforming protein 21) in single spines undergoing structural plasticity and long-term potentiation<sup>12,13</sup>. CaMKII activation is restricted to spines, and decays rapidly with a time constant of about ten seconds<sup>13</sup>. In contrast, HRas activity spreads from the stimulated spines along dendrites and into surrounding spines over about 10  $\mu\text{m}$  (ref. 12). However, to achieve long-lasting, spine-specific plasticity, there should also exist signalling pathways that relay compartmentalized signalling on the timescale of minutes to hours. Rho GTPases may constitute such signalling, because they are important in regulating the actin cytoskeleton<sup>3,16</sup>, which is essential for spine-specific, long-term structural and functional plasticity<sup>14,17</sup>.

To measure the activation of Rho GTPases in single dendritic spines, we developed fluorescence resonance energy transfer (FRET)-based sensors optimized for imaging under 2pFLIM using a design similar to a previously developed HRas sensor<sup>11</sup>. The RhoA/Cdc42 sensors consist of two components: RhoA/Cdc42 tagged with monomeric enhanced green fluorescent protein (mEGFP) and their binding partner, Rho GTPase binding domain (RBD) of Rhotekin/Pak3, doubly tagged with mCherry (mCherry-RBD-mCherry) (Supplementary note). When mEGFP–Rho GTPase is activated, mCherry–RBD–mCherry binds to mEGFP–RhoA/Cdc42, causing FRET between mEGFP and mCherry (Supplementary Figs 1 and 2). These sensors were verified to be specific and sensitive under 2pFLIM (Supplementary note).

Using these sensors, we measured the activity of RhoA and Cdc42 during spine structural plasticity associated with long-term potentiation



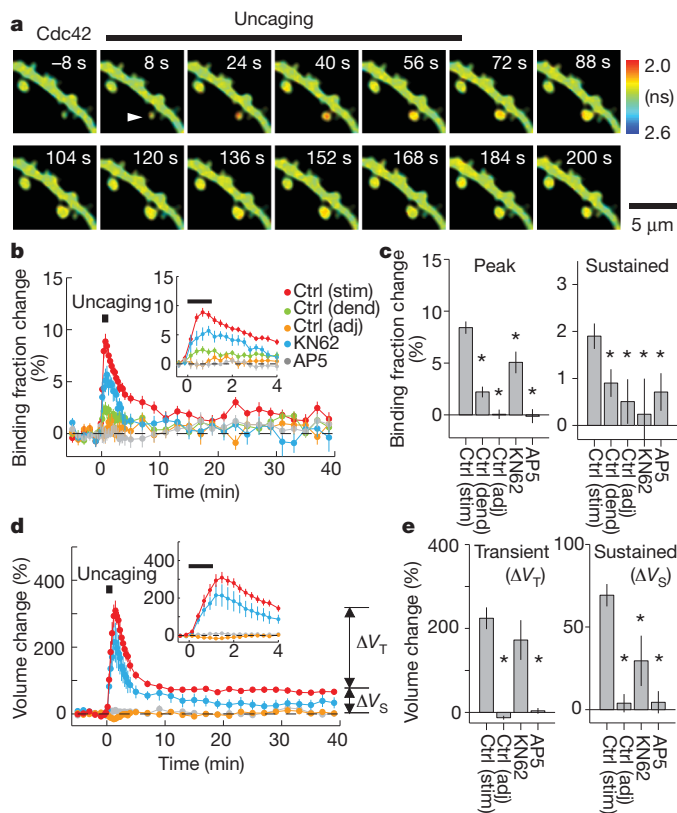
**Figure 1 | Spatiotemporal dynamics of RhoA activation during long-term structural plasticity induced in single spines** **a**, Fluorescence lifetime images of RhoA activation during spine structural plasticity induced by two-photon glutamate uncaging. Arrowhead indicates the stimulated spine. Warmer colours indicate shorter lifetimes and higher RhoA activity. Scale bar, 5  $\mu\text{m}$ . **b**, Time course of RhoA activation measured as a change in the fraction of mEGFP–RhoA bound to mCherry–RBD–mCherry in stimulated spines (stim), the dendritic shaft beside the stimulated spines (dend; within 1  $\mu\text{m}$ ), and adjacent spines (adj; between 3–5  $\mu\text{m}$  of the stimulated spines). Data using pharmacological inhibitors (Ctrl, control condition; KN62, CaMKII inhibitor; AP5, NMDA receptor inhibitor) are also shown. The inset to **b** shows a closer view of the first 4 min. The numbers of samples (spines/neurons) are 35/29 for stimulated spines and dendrites, 29/26 for adjacent spines, 16/10 for KN62 and 8/5 AP5. Error bars are s.e.m. **c**, Transient (averaged over 16–64 s) and sustained (averaged over 20–38 min) RhoA activation. Stars denote statistically significant difference ( $<0.05$ ) from the value in the stimulated spines under the control condition. Wilcoxon signed-rank test was used for dendrites and adjacent spines, and analysis of variance (ANOVA) followed by post-hoc tests using the least significant difference was used for experiments with pharmacological inhibitors. **d**, Averaged time course of spine volume change in the same experiments as in **b**. The inset to **d** shows a closer view of the first 4 min. **e**, Transient (volume change averaged over 1.5–2 min subtracted by that over 20–38 min) and sustained volume change (volume change averaged over 20–38 min).

<sup>1</sup>Department of Neurobiology, <sup>2</sup>Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA.

(Figs 1, 2 and 3). Pyramidal neurons in the CA1 region of cultured hippocampal slices were ballistically<sup>18</sup> transfected with the RhoA or Cdc42 sensor, and the FRET signal was imaged under 2pFLIM. The spine volume was monitored using the red fluorescence of mCherry-RBD-mCherry (Supplementary Fig. 3)<sup>12</sup>. To induce structural plasticity in a single dendritic spine, we applied a low-frequency train of two-photon glutamate uncaging pulses (30 pulses at 0.5 Hz) to the spine in zero extracellular  $Mg^{2+}$  (refs 13, 14 and 19). The spine volume increased rapidly by about 300% following glutamate uncaging (transient phase) and relaxed to an elevated level of 70–80% for more than 30 min (sustained phase) (Figs 1d and 2d)<sup>12–14</sup>. The time course of spine enlargement in neurons expressing the FRET sensor was similar to that in neurons expressing only EGFP (Fig. 4)<sup>14</sup>, suggesting that the overexpression of FRET sensors causes almost no effects on spine structural plasticity (Supplementary note).

Under basal conditions, there was no correlation between the activity of Rho GTPases and spine volume (Supplementary Fig. 4). When spine structural plasticity was induced, both RhoA (Fig. 1a and b) and Cdc42 (Fig. 2a and b) were activated rapidly within about 30 s in the stimulated spines. The activation decayed over about 5 min, followed by sustained activity lasting more than 30 min. RhoA activation spread into the dendrites over several micrometres (Figs 1a–c, and 3a and b), and invaded surrounding spines to a small extent (~25% of the stimulated spines). In contrast, Cdc42 activation was restricted to the stimulated spines (Figs 2a–c, and 3c and d). For both RhoA and Cdc42, the gradient at the spine necks was maintained for more than about 30 min (Figs 1b and c, and 2b and c).

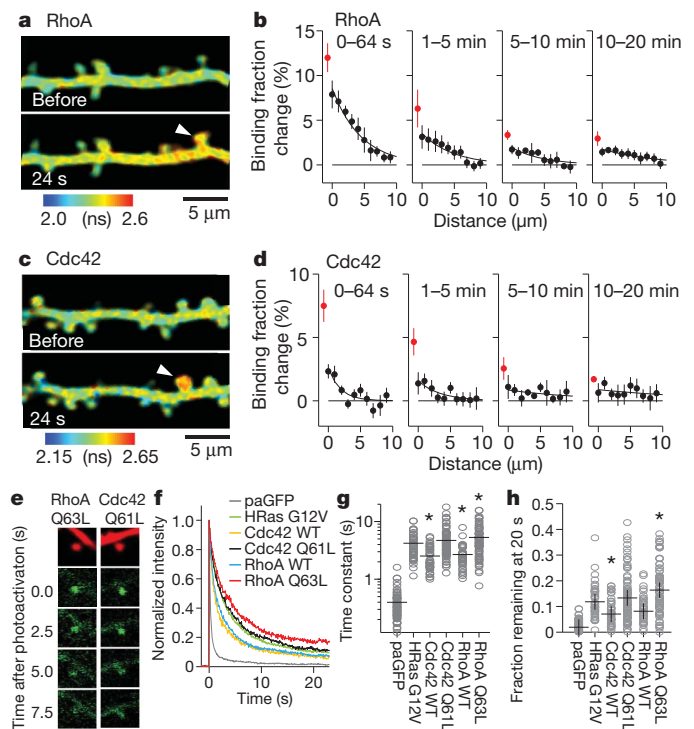
Next, we pharmacologically identified the signalling pathways that activate RhoA and Cdc42. Inhibition of *N*-methyl-D-aspartic acid (NMDA) receptors with 2-amino-5-phosphonopentanoic acid (AP5, 50  $\mu$ M) abolished activation of RhoA (Fig. 1b and c) and Cdc42 (Fig. 2b



**Figure 2 | Spatiotemporal dynamics of Cdc42 activation during long-term structural plasticity induced in single spines.** The same experiments and analyses as in Fig. 1 but measuring Cdc42 activity instead of RhoA activity. The numbers of samples (spines/neurons) are 33/28 for stimulated spines and dendrite, 33/28 for adjacent spines, 11/6 for KN62 and 12/8 for AP5.

and c) as well as spine enlargement (Figs 1d and e, and 2d and e)<sup>14</sup>, indicating that RhoA and Cdc42 are activated by  $Ca^{2+}$  through NMDA receptors. The CaMKII inhibitor KN62 is known to strongly inhibit sustained spine enlargement, but has significantly less of an effect on transient spine enlargement (Figs 1d and e, and 2d and e)<sup>13,14</sup>. KN62 (10  $\mu$ M) partially inhibited RhoA and Cdc42 activation during the transient phase, and more strongly during the sustained phase. (Figs 1b and c, and 2b and c). Expression of autocamtide CaMKII inhibitor peptide 2 (AIP2) also inhibited spine volume change and Rho GTPase activation in a similar manner (Supplementary Fig. 5). These results suggest that RhoA and Cdc42 are downstream of CaMKII.

We next characterized the spatial profile of RhoA and Cdc42 activities along dendrites as a function of the distance from the stimulated spines (Fig. 3). RhoA activity showed a relatively small gradient between the stimulated spines and dendrites, and spread along the dendrites. The length constant of spread along the dendrite was 4.5  $\mu$ m (Fig. 3a and b), a value similar to that for HRas (about 10  $\mu$ m)<sup>12</sup>. In contrast, Cdc42 activation was restricted to the stimulated spines (Fig. 3c and d), showing a spatial pattern similar to that of CaMKII<sup>13</sup>. A small fraction of Cdc42 activation spread into the



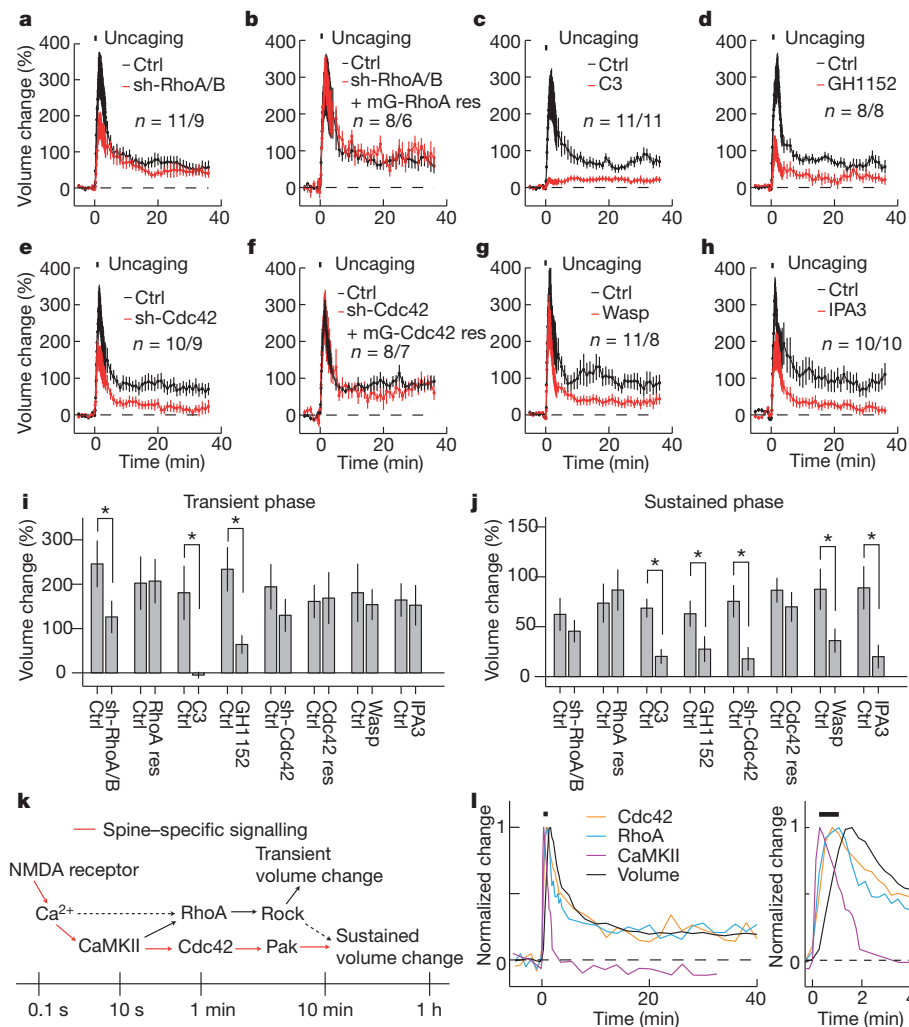
**Figure 3 | Spatial profile of RhoA and Cdc42 activities.** a, Fluorescence lifetime images of RhoA activity before and after glutamate uncaging. Arrowheads in a and c indicate the stimulated spine. b, Averaged spatial profile of RhoA activation. Red circles indicate the activity in the stimulated spine, and black circles indicate the activity in the dendrite, plotted as a function of the distance along the dendrite from the simulated spine. The number of samples (dendrites/neurons) is 20/18. c, Fluorescence lifetime images of Cdc42 activity. d, Averaged spatial profile of Cdc42 activation. The number of samples (dendrites/neurons) is 30/26. e, The fluorescence images of paGFP–RhoA (left) and paGFP–Cdc42 (right) after spine-head photoactivation (green, paGFP–Rho GTPases; red, tandem mCherry). f, Averaged timecourse of fluorescence decay in spines after photoactivation of paGFP-tagged proteins in the spines. The fluorescence intensity was normalized to the peak. The numbers of samples (spines/neurons) are 63/6 for paGFP, 38/4 for paGFP–HRas (G12V), 41/5 for paGFP–Cdc42 (WT), 83/9 for paGFP–Cdc42 (Q61L), 40/4 for paGFP–RhoA (WT) and 79/10 for paGFP–RhoA (Q63L). HRas (G12V), RhoA (Q63L) and Cdc42 (Q61L) are constitutively active mutants. g, h, Decay time constants (g) and the fraction remaining at 20 s (h) of paGFP fluorescence in the photoactivated spines. Horizontal bars indicate the means.

dendrite and decayed sharply with a length constant of around  $1.9\ \mu\text{m}$  (Fig. 3c and d). These experiments were performed at room temperature ( $25\text{--}27^\circ\text{C}$ ), but similar results were also obtained at near-physiological temperature ( $32\text{--}34^\circ\text{C}$ ; Supplementary Figs 6 and 7).

To test whether the difference in the degree of the compartmentalization of RhoA and Cdc42 is due to a difference in their mobility<sup>12,13</sup>, we measured spine–dendrite diffusion coupling using photoactivatable GFP (paGFP)<sup>20</sup> fused to Rho GTPases. Following photoactivation of paGFP in a spine, the fluorescence intensity in the spine decayed owing to the diffusion of paGFP–Rho GTPases out of the spine with a time constant of about 3 s for the wild type and about 5 s for their constitutively active mutants (Fig. 3e–h). These values are about ten times larger than the decay time constant of cytosolic paGFP ( $\sim 0.4$  s) and similar to that of a constitutively active HRas mutant ( $\sim 5$  s) (Fig. 3e–h)<sup>12</sup>. The difference between wild-type Rho GTPases and constitutively active mutants presumably reflects the difference in the fraction of the protein

bound to the plasma membrane, given that active Rho GTPases are localized on the plasma membrane<sup>21</sup>. There was only a small fraction (10–20%) of fluorescence remaining at 20 s after photoactivation (Fig. 3e–h), suggesting that no major immobile fraction of RhoA or Cdc42 exists in spines. Thus, Cdc42 is as mobile as RhoA and HRas, yet only Cdc42 shows the compartmentalized activity.

Next, to elucidate the roles of Rho GTPase activation in spine structural plasticity, we measured the spine volume change under the inhibition of Rho or Cdc42 signalling (Fig. 4a–j). Downregulation of RhoA and RhoB with short-hairpin RNA (shRNA) decreased the transient volume change, but did not appreciably affect the sustained volume change (Fig. 4a, i and j). In contrast, shRNA against Cdc42 decreased the sustained volume change, but not the transient volume change (Fig. 4e, i and j). The phenotypes caused by these shRNAs were rescued by co-expressing shRNA-resistant mEGFP–Rho GTPases (Fig. 4b, f, i and j), indicating that the effect of the shRNAs is specific



**Figure 4 | The effect of Rho GTPase inhibition for structural plasticity of spine head enlargement.** **a–h**, Averaged time course of spine volume change in stimulated spines in neurons under manipulations of Rho GTPase signalling. Red traces: neurons were transfected with shRNAs against RhoA and RhoB (sh-RhoA/B) and mEGFP (**a**), sh-RhoA/B, mEGFP–shRNA resistant RhoA (mEGFP–RhoA res) and tandem mCherry (**b**), mCherry–C3 transferase (C3) and mEGFP (**c**), mEGFP (**d**), shRNA against Cdc42 (sh-Cdc42) and mEGFP (**e**), sh-Cdc42, mEGFP–shRNA resistant Cdc42 (mEGFP–Cdc42 res) and tandem mCherry (**f**) or mCherry–Wasp(210–321) (Wasp) and mEGFP (**g**). Black traces: paired control experiments were performed in the same batch of slices using a scrambled shRNA instead of targeted shRNAs (**a**, **e**), mEGFP alone (**b**, **f**) or mCherry instead of C3 transferase and Wasp (**c**, **g**). Pharmacology experiments (**d**, **h**) were performed before (paired control,

black) and 30–40 min (red) after applying drugs to the bath. Fluorescence intensity of mEGFP (**a**, **c**, **d**, **e**, **g**, **h**) or tandem mCherry (**b**, **f**) was used to measure the spine volume change. The numbers of samples (spines/neurons) are indicated in the figures (same numbers for control and experiment groups). **i**, Transient volume change (volume change averaged over 1.5–2 min subtracted by that averaged over 20–36 min). Stars denote statistical significance ( $P < 0.05$ , paired  $t$ -test). **j**, Sustained volume change (volume change averaged over 20–36 min). **k**, A model of Cdc42 and RhoA activation. **l**, Superimposed time courses of spine volume change and activation of RhoA (Fig. 1b), Cdc42 (Fig. 2b) and CaMKII<sup>13</sup> in spines undergoing structural plasticity. The time courses were normalized to the peak. The right-hand panel shows a closer view of the first 4 min.



and the mEGFP–RhoA and mEGFP–Cdc42 used in the FRET sensors are functional as endogenous proteins. Because downregulation of proteins with shRNA is partial (Supplementary Fig. 8) and requires a relatively long time (4 days), we also inhibited Rho and Cdc42 signalling by expressing mCherry–C3 transferase, a Rho inhibitor<sup>22</sup>, and the Cdc42 binding domain of Wasp (221–321) tagged with mCherry (Wasp)<sup>23</sup>, respectively, for shorter time (24 h). Rho inhibition with C3 transferase inhibited both the transient and the sustained phases (Fig. 4c, i and j), showing stronger effects than shRNA (Fig. 4a, i and j). Cdc42 inhibition with Wasp inhibited the sustained phase but not the transient phase (Fig. 4g, i and j), consistent with the shRNA result (Fig. 4e, i and j). Thus, our data suggest that Rho signalling is required for the transient phase and probably the sustained phase of spine enlargement, whereas Cdc42 signalling is required for the sustained phase. Neither C3 transferase nor Wasp affected CaMKII activation (Supplementary Fig. 9), indicating that there is no feedback signalling from Rho and Cdc42 to upstream  $\text{Ca}^{2+}$  and CaMKII. C3 transferase and Wasp also inhibited synaptic potentiation induced by pairing postsynaptic depolarization (0 mV) and two-photon glutamate uncaging (Supplementary Fig. 10)<sup>13,14</sup>, suggesting that Rho and Cdc42 are important for the functional plasticity as well as the structural plasticity of spines.

Among known effectors of Rho and Cdc42, Rock and Pak are two kinases that can be activated respectively by these GTPases<sup>24–26</sup>. We tested whether they are required for structural plasticity through acute (30–40 min) application of specific pharmacological inhibitors. Inhibition of Rock with Glycyl-H1152 (2  $\mu\text{M}$ )<sup>27</sup> suppressed both transient and sustained volume change (Fig. 4d, i and j), similarly to the Rho inhibitor C3 transferase (Fig. 4c). In contrast, inhibition of Pak with inhibitor targeting Pak1 activation-3 (IPA3) (100  $\mu\text{M}$ )<sup>28</sup> decreased the sustained volume change selectively without changing the transient volume change (Fig. 4h, i and j), similarly to inhibition of Cdc42 signalling (Fig. 4e and g). Taken together with the results from Rho/Cdc42 inhibition (Fig. 4a–j), our data implies that the Rho–Rock pathway is required for both the transient and sustained phases, whereas the Cdc42–Pak pathway is required for the sustained phase of the structural plasticity but not for the transient phase (Fig. 4k).

In this study, we visualized RhoA and Cdc42 activation in single dendritic spines undergoing structural plasticity associated with long-term potentiation<sup>12–15,19</sup>. The time course of their activation was similar to that of the volume change: rapid activation was followed by persistent activation lasting more than 30 min (Fig. 4l). As expected from its high mobility (Fig. 3), RhoA spread into the dendrite upon activation (Fig. 1a–c)<sup>13</sup>. However, the activity invasion into adjacent spines was relatively small (25% of the stimulated spines, Fig. 1b and c) and was not sufficient to produce plasticity (Fig. 1d and e). In contrast with the diffusive pattern of RhoA activity, Cdc42 activity was restricted to the stimulated spines (Figs 2 and 3). The compartmentalization of Cdc42 activity is not due to the limited diffusion of active Cdc42, because active Cdc42 is as mobile as RhoA and HRas (Fig. 3e–h). Given that the high spatial gradient of Cdc42 between the stimulated spines and dendrite was maintained for more than 30 min (Fig. 2b and c, and 3d), Cdc42 must be continuously activated at the stimulated spines during plasticity, and inactivated immediately after diffusing out of the spines. The short length constant of Cdc42 in the dendrites (1.9  $\mu\text{m}$ , Fig. 3) also supports the fast inactivation of Cdc42 in the dendrite. The inactivation time constant  $\tau$  is related to the length constant  $L$  and the diffusion constant  $D \approx 0.6 \mu\text{m}^2$  (ref. 12) as follows:  $\tau = L^2/D$ , and so  $\tau$  is about 6 s for Cdc42 and about 30 s for RhoA, compared to 200–300 s for HRas<sup>12</sup>.

Our results further indicated that RhoA and Cdc42 activation is CaMKII-dependent, and activation lasts for more than 30 min (Figs 1 and 2). The previous imaging study suggested that CaMKII activity decays with a time constant of about 10 s (Fig. 4k and l)<sup>13</sup>, thereby integrating NMDA-receptor-evoked  $\text{Ca}^{2+}$  transients that last for about 0.1 s (refs 13 and 29). Localized, persistent activation of RhoA and

Cdc42, which peaks between CaMKII activation and the volume change (Fig. 4l), relays the transient CaMKII signalling<sup>13</sup> to long-term spine structural plasticity (Fig. 4k). In particular, because both CaMKII<sup>13,14</sup> and Cdc42 exhibit spine specific activation and are required for the maintenance of plasticity (Fig. 4e, g, i and j), the NMDA receptor–CaMKII–Cdc42–Pak pathway (red line in Fig. 4k) constitutes the spine-specific signal transduction spanning the timescale from less than one second to more than half an hour to cause sustained structural and functional spine plasticity (Fig. 4k and l).

## METHODS SUMMARY

Hippocampal cultured slices were prepared from postnatal day 6–7 rats as described<sup>30</sup>. Neurons were sparsely transfected with Rho GTPase FRET sensors using ballistic gene transfer<sup>18</sup> at days *in vitro* 10–14, and imaged 2–4 days after transfection. Rho–GTPase activity was measured using 2pFLIM (green) and spine volume change was monitored by measuring the fluorescence intensity of mCherry–RBD–mCherry (red) in spines using normal two-photon microscopy (Supplementary Fig. 3)<sup>12,13</sup>. Most of the imaging experiments were performed at room temperature (25–27 °C).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 11 March 2010; accepted 13 January 2011.**

**Published online 20 March 2011.**

1. Tashiro, A. & Yuste, R. Regulation of dendritic spine motility and stability by Rac1 and Rho kinase: evidence for two forms of spine motility. *Mol. Cell. Neurosci.* **26**, 429–440 (2004).
2. Luo, L. Rho GTPases in neuronal morphogenesis. *Nature Rev. Neurosci.* **1**, 173–180 (2000).
3. Saneyoshi, T., Fortin, D. A. & Soderling, T. R. Regulation of spine and synapse formation by activity-dependent intracellular signaling pathways. *Curr. Opin. Neurobiol.* **20**, 108–115 (2009).
4. Fortin, D. A. *et al.* Long-term potentiation-dependent spine enlargement requires synaptic  $\text{Ca}^{2+}$ -permeable AMPA receptors recruited by CaM-kinase I. *J. Neurosci.* **30**, 11565–11575 (2010).
5. O’Kane, E. M., Stone, T. W. & Morris, B. J. Activation of Rho GTPases by synaptic transmission in the hippocampus. *J. Neurochem.* **87**, 1309–1312 (2003).
6. Rex, C. S. *et al.* Different Rho GTPase-dependent signaling pathways initiate sequential steps in the consolidation of long-term potentiation. *J. Cell Biol.* **186**, 85–97 (2009).
7. Asrar, S. *et al.* Regulation of hippocampal long-term potentiation by p21-activated protein kinase 1 (PAK1). *Neuropharmacology* **56**, 73–80 (2009).
8. Wang, H. G. *et al.* Presynaptic and postsynaptic roles of NO, cGK, and RhoA in long-lasting potentiation and aggregation of synaptic proteins. *Neuron* **45**, 389–403 (2005).
9. Nadif Kasri, N. & Van Aelst, L. Rho-linked genes and neurological disorders. *Pflügers Arch.* **455**, 787–797 (2008).
10. Hotulainen, P. & Hoogenraad, C. C. Actin in dendritic spines: connecting dynamics to function. *J. Cell Biol.* **189**, 619–629 (2010).
11. Yasuda, R. *et al.* Supersensitive Ras activation in dendrites and spines revealed by two-photon fluorescence lifetime imaging. *Nature Neurosci.* **9**, 283–291 (2006).
12. Harvey, C. D., Yasuda, R., Zhong, H. & Svoboda, K. The spread of Ras activity triggered by activation of a single dendritic spine. *Science* **321**, 136–140 (2008).
13. Lee, S. J., Escobedo-Lozoya, Y., Szatmari, E. M. & Yasuda, R. Activation of CaMKII in single dendritic spines during long-term potentiation. *Nature* **458**, 299–304 (2009).
14. Matsuzaki, M., Honkura, N., Ellis-Davies, G. C. & Kasai, H. Structural basis of long-term potentiation in single dendritic spines. *Nature* **429**, 761–766 (2004).
15. Honkura, N., Matsuzaki, M., Noguchi, J., Ellis-Davies, G. C. & Kasai, H. The subspine organization of actin fibers regulates the structure and plasticity of dendritic spines. *Neuron* **57**, 719–729 (2008).
16. Heasman, S. J. & Ridley, A. J. Mammalian Rho GTPases: new insights into their functions from *in vivo* studies. *Nature Rev. Mol. Cell Biol.* **9**, 690–701 (2008).
17. Okamoto, K., Nagai, T., Miyawaki, A. & Hayashi, Y. Rapid and persistent modulation of actin dynamics regulates postsynaptic reorganization underlying bidirectional plasticity. *Nature Neurosci.* **7**, 1104–1112 (2004).
18. McAllister, A. K. Biolistic transfection of neurons. *Sci. STKE* **2000**, pl1 (2000).
19. Harvey, C. D. & Svoboda, K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. *Nature* **450**, 1195–1200 (2007).
20. Patterson, G. H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297**, 1873–1877 (2002).
21. Michaelson, D. *et al.* Differential localization of Rho GTPases in live cells: regulation by hypervariable regions and RhoGDI binding. *J. Cell Biol.* **152**, 111–126 (2001).
22. Wilde, C., Genth, H., Aktories, K. & Just, I. Recognition of RhoA by *Clostridium botulinum* C3 exoenzyme. *J. Biol. Chem.* **275**, 16478–16483 (2000).
23. Elliot-Smith, A. E., Mott, H. R., Lowe, P. N., Laue, E. D. & Owen, D. Specificity determinants on Cdc42 for binding its effector protein ACK. *Biochemistry* **44**, 12373–12383 (2005).

24. Nikolić, M. The Pak1 kinase: an important regulator of neuronal morphology and function in the developing forebrain. *Mol. Neurobiol.* **37**, 187–202 (2008).
25. Kreis, P. *et al.* The p21-activated kinase 3 implicated in mental retardation regulates spine morphogenesis through a Cdc42-dependent pathway. *J. Biol. Chem.* **282**, 21497–21506 (2007).
26. Iden, S. & Collard, J. G. Crosstalk between small GTPases and polarity proteins in cell polarization. *Nature Rev. Mol. Cell Biol.* **9**, 846–859 (2008).
27. Tamura, M. *et al.* Development of specific Rho-kinase inhibitors and their clinical application. *Biochim. Biophys. Acta* **1754**, 245–252 (2005).
28. Deacon, S. W. *et al.* An isoform-selective, small-molecule inhibitor targets the autoregulatory mechanism of p21-activated kinase. *Chem. Biol.* **15**, 322–331 (2008).
29. Sabatini, B. L., Oertner, T. G. & Svoboda, K. The life cycle of Ca(2+) ions in dendritic spines. *Neuron* **33**, 439–452 (2002).
30. Stoppini, L., Buchs, P. A. & Müller, D. A simple method for organotypic cultures of nervous tissue. *J. Neurosci. Methods* **37**, 173–182 (1991).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Hahn, G. Bokock, A. Aplin, S. Soderling, M. Matsuda and Y. Hayashi for plasmids and reagents, S.-J. Lee for CaMKII activation data, S. Soderling, S. Raghavachari, L. van Aelst, Y. Hayashi, H. Kasai and M. Ehlers for discussion, and H. Hedrick and M. Patterson for comments on the manuscript. We also thank A. Wan for preparing cultured slices and D. Kloetzer for laboratory management. This study was funded by the Howard Hughes Medical Institute, the National Institute of Mental Health, the National Institute of Neurological Disorders and Stroke, the National Institute of Drug Abuse, the Alzheimer's Association and the Japan Society for the Promotion of Science (H.M.).

**Author Contributions** H.M. and R.Y. designed the experiments. H.M. performed the experiments and data analysis. H.W. performed electrophysiological experiments. H.M. and R.Y. wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to R.Y. ([yasuda@neuro.duke.edu](mailto:yasuda@neuro.duke.edu)).

## METHODS

**Plasmids.** Plasmids containing Dbl/p50RhoGAP, Rhotekin(8–89), C3 transferase, RhoA/Cdc42/Pak1(65–118), and Wasp were the kind gifts of K. Hahn, G. Bokock, A. Apelin, M. Matsuda and S. Soderling, respectively. Pak3(60–113) was prepared by introducing mutations M99I and S115L into Pak1(65–118). mEGFP–RhoA and mEGFP–Cdc42 were prepared by inserting human RhoA and Cdc42 coding sequences into the pEGFP–C1 vector (Clontech) containing the A206K monomeric mutation in EGFP<sup>31</sup>. Because the transcription of human RhoA is terminated at the codon 349–353 (AATAA) in *Escherichia coli*, we introduced a silent mutation from AATAA to AACAA at the site, which does not change the amino acid sequence. This sequence was used for all experiments. The linker between mEGFP and Rho GTPases (RhoA and Cdc42) is SGLRSRG. Tandem mCherry<sup>32</sup> was prepared by replacing the EGFP of the pEGFP–N3 vector with two mCherry. mCherry–Rhotekin(8–89)–mCherry and mCherry–Pak3(60–113)–mCherry were prepared by inserting Rhotekin(8–89) and Pak3(60–113) into tandem mCherry, respectively. The linkers between Rhotekin(8–89) and mCherry are SGLRSRG for the amino terminus and VDVDTAGPGSG for the carboxy terminus. The linkers between Pak3(60–113) and mCherry are SGLRSRG for the N terminus and GSG for the C terminus. Photoactivatable GFP (paGFP)<sup>20</sup>–Rho GTPases were prepared by replacing the mEGFP of the mEGFP–Rho GTPases with paGFP (A206K). Mutations were introduced using a Site-Directed Mutagenesis kit (Stratagene). shRNAs were prepared using the pSuper vector (Oligoengine) with kanamycin resistance gene. The following target sequences were used for shRNA (5′–3′): GGGCAAGAGGATTATGACA for Cdc42 (rat and human), GAAGGATCTTCGGAATGAT for RhoA (mouse, rat and human), CATCTTGTTGGCCAACAAA for RhoB (rat) and GTGTGAAGTATCTGTACG for control. shRNA-resistant RhoA and Cdc42 were prepared by introducing three silent mutations in the targeted sequences. mCherry–C3 transferase (34–end) and mCherry–Wasp(210–321) were prepared into the pEGFP–C1 vector without EGFP.

**Proteins.** Polyhistidine-tagged mEGFP, mCherry, super-folder GFP (sfGFP)–Rho GTPases<sup>33</sup>, mCherry–RBDs and their mutants were cloned into the pRSET bacterial expression vector (Invitrogen). Proteins were overexpressed in *E. coli* (DH5α), purified with a Ni<sup>2+</sup>-nitrilotriacetate column (HiTrap, GE Healthcare), and desalted with a desalting column (PD10, GE Healthcare) equilibrated with phosphate buffered saline. The concentration of the purified protein was measured by the absorbance of the fluorophore (mEGFP,  $A_{489\text{ nm}} = 56,000\text{ cm}^{-1}\text{ M}^{-1}$  (ref. 13); sfGFP,  $A_{489\text{ nm}} = 83,000\text{ cm}^{-1}\text{ M}^{-1}$  (ref. 33); mCherry,  $A_{587\text{ nm}} = 72,000\text{ cm}^{-1}\text{ M}^{-1}$  (ref. 32)) or Bradford assay.

**Preparation.** Hippocampal slices were prepared from postnatal day 6–7 rats, as described<sup>30</sup>, in accordance with the animal care and use guidelines of Duke University Medical Center. After 1 to 2 weeks in culture, CA1 pyramidal neurons were transfected with ballistic gene transfer<sup>34</sup> using gold beads (8–12 mg) coated with plasmids containing 30 μg of total complementary DNA (donor:acceptor = 1:1), and imaged 2–4 days after transfection.

HeLa and Rat1 cells were cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal calf serum at 37 °C in 5% CO<sub>2</sub> and transfected using Lipofectamine (Invitrogen).

**Two-photon fluorescence lifetime imaging.** Details of FRET imaging using a custom-built two-photon fluorescence lifetime imaging microscope have been described previously<sup>35,36</sup>. mEGFP and mCherry were simultaneously excited with a Ti:sapphire laser (Maitai, Spectraphysics) tuned at a wavelength of 920 nm. The fluorescence was collected by an objective (60×, numerical aperture 0.9, Olympus), divided with a dichroic mirror (565 nm) and detected with two separated photoelectron multiplier tubes placed after wavelength filters (Chroma, HQ510/70–2p for green and HQ620/90–2p for red). For fluorescence lifetime imaging in the green channel, a photoelectron multiplier tube with low transfer time spread (H7422–40p; Hamamatsu) was used. A wide-aperture photoelectron multiplier tube (R3896; Hamamatsu) was used for the red channel. Fluorescence lifetime images were obtained using a time-correlated single photon counting board (SPC-140; Becker and Hickl) controlled with custom software<sup>31</sup>. The red signal was acquired using a separate data acquisition board (PCI-6110) and Scanimage software<sup>37</sup>.

**Two-photon glutamate uncaging.** A second Ti:sapphire laser tuned at a wavelength of 720 nm was used to uncage 4-methoxy-7-nitroindolyl-caged-L-glutamate (MNI-caged glutamate) in extracellular solution with a train of 4–6-ms, 8-mW pulses (30 times at 0.5 Hz) near a spine of interest. Experiments were performed in Mg<sup>2+</sup>-free artificial cerebral spinal fluid (127 mM NaCl, 2.5 mM KCl, 4 mM CaCl<sub>2</sub>, 25 mM NaHCO<sub>3</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub> and 25 mM glucose) containing 1 μM tetrodotoxin and 2 mM MNI-caged L-glutamate aerated with 95% O<sub>2</sub> and 5% CO<sub>2</sub> at 25–27 °C, as described previously<sup>13</sup>. In Supplementary Figs 6 and 7, neurons were maintained at 32–34 °C using a temperature controller (TC324B, SW-10/6 and SH-27B, Warner Instruments).

**Two-photon photoactivation of photoactivatable GFP.** Two-photon images of paGFP tagged proteins were acquired every 64 ms using a Ti:sapphire laser tuned at a wavelength of 940 nm. For photoactivation of paGFP, the uncaging laser tuned at a wavelength of 800 nm was used to apply a pulse of 8 mW with 6–10 ms duration at a spine head. To determine the decay time constant  $\tau$  and the immobile fraction  $f_{\text{im}}$ , the paGFP fluorescence  $F$  was fitted with an exponential function,  $F(t) = F_0 \exp(-t/\tau) + f_{\text{im}}$ , where  $F_0$  is the fluorescence intensity at  $t = 0$ .

**2pFLIM data analyses.** To obtain the mEGFP fluorescence lifetime, we summed over all pixels in an image of a cell expressing mEGFP–Rho GTPases, and fitted a fluorescence lifetime curve with a single exponential function convolved with the Gaussian pulse response function:

$$F(t) = F_0 H(t, t_0, \tau_D, \tau_G) \quad (1)$$

where  $F_0$  is the constant, and

$$H(t, t_0, \tau_D, \tau_G) = \frac{1}{2} \exp\left(\frac{\tau_G^2}{2\tau_D} - \frac{t-t_0}{\tau_D}\right) \text{erf}\left(\frac{\tau_G^2 - \tau_D(t-t_0)}{\sqrt{2\tau_D\tau_G}}\right) \quad (2)$$

in which  $\tau_D$  is the fluorescence lifetime of the free donor (mEGFP–Rho GTPase),  $\tau_G$  is the width of the Gaussian pulse response function,  $F_0$  is the peak fluorescence before convolution and  $t_0$  is the time offset, and erf is the error function. To measure the fraction of donor bound to acceptor, we summed all pixels over a whole image and fitted a fluorescence lifetime curve with a double exponential function convolved with the Gaussian pulse response function:

$$F(t) = F_0 [P_D H(t, t_0, \tau_D, \tau_G) + P_{AD} H(t, t_0, \tau_{AD}, \tau_G)] \quad (3)$$

where  $\tau_{AD}$  is the fluorescence lifetime of donor bound with acceptor, and  $P_D$  and  $P_{AD}$  are the fractions of free donor and donor bound with acceptor, respectively. We fixed  $\tau_D$  to the fluorescence lifetime obtained from free mEGFP–Rho GTPase (2.59 ns). To generate the fluorescence lifetime image, we calculated the mean photon arrival time,  $\langle t \rangle$ , in each pixel as<sup>38</sup>:

$$\langle t \rangle = \int t F(t) dt / \int F(t) dt$$

The mean photon arrival time is then related to the mean fluorescence lifetime  $\langle \tau \rangle$  by an offset arrival time  $t_0$ , which is obtained by fitting the whole image<sup>35</sup>:

$$\langle \tau \rangle = \langle t \rangle - t_0$$

For small regions of interest in an image (spines or dendrites), we calculated the binding fraction ( $P_{AD}$ ) as<sup>35</sup>:

$$P_{AD} = \tau_D(\tau_D - \langle \tau \rangle)(\tau_D - \tau_{AD})^{-1}(\tau_D + \tau_{AD} - \langle \tau \rangle)^{-1} \quad (4)$$

**Overexpression level of Rho GTPase sensors in neurons.** The concentration of mEGFP–Rho GTPase and mCherry–RBD–mCherry in neurons was estimated by measuring fluorescence intensity of mEGFP and mCherry in thick apical dendrites under two-photon microscopy relative to that of purified, polyhistidine-tagged mEGFP (1 μM) and mCherry (10 μM), respectively<sup>39</sup>.

**Measurements of the affinity between Rho GTPases and RBDs.** Purified sfGFP–Rho GTPases (RhoA and Cdc42) were loaded with GppNHp (2′,3′-O-N-methyl anthraniloyl-GppNHp) and GDP by incubating in the presence of tenfold molar excess of GppNHp and GDP in MgCl<sub>2</sub>-free phosphate buffered saline containing 1 mM EDTA for 10 min, respectively. The reaction was terminated by adding 10 mM MgCl<sub>2</sub> (ref. 40). sfGFP–Rho GTPases and mCherry–RBD were mixed and incubated at room temperature for 20 min. FRET between sfGFP and mCherry was measured under 2pFLIM, and the fraction of sfGFP–Rho GTPases bound to mCherry–RBD was calculated by fitting the fluorescence lifetime curve with a double exponential function (equation (3)). The dissociation constant was obtained by fitting the relationship between the binding fraction and the concentration of mCherry–RBD with a Michaelis–Menten function (Supplementary Fig. 1).

**Estimation of endogenous Rho GTPase concentration.** We determined the concentrations of CaMKIIα, RhoA, and Cdc42 in the CA1 region of hippocampal slice culture by semiquantitative western blotting (Supplementary Fig. 15). First, the CA1 regions from ten slices were collected and weighed. The series of purified CaMKIIα (4 μM, 10 μM, 20 μM, 30 μM, 40 μM), RhoA (0.1 μM, 0.2 μM, 0.5 μM, 1.0 μM, 1.5 μM), or Cdc42 (0.1 μM, 0.2 μM, 0.5 μM, 1.0 μM, 1.5 μM) were prepared to the same weights as the CA1 tissue. The CA1 tissue and purified proteins were dissolved in SDS sample buffer, and analysed by western blotting. The following antibodies were used: anti-CaMKII (EP1829Y; Abcam); anti-RhoA (26C4; Santa Cruz Biotechnology); anti-Cdc42 (BD44; BD Transduction Laboratories); goat anti-mouse (Zymax). Chemiluminescence signals were detected using a Storm image acquisition system (Molecular Dynamics), and analysed digitally using ImageJ software. By comparing the band intensities of the purified proteins



to that of lysates from the CA1 tissue, we calculated the concentration of the respective proteins in the CA1 tissue (Supplementary Fig. 6). The concentration was estimated to be 23.4  $\mu$ M for CaMKII $\alpha$ , 0.38  $\mu$ M for RhoA and 0.25  $\mu$ M for Cdc42. The estimation of CaMKII $\alpha$  concentration is consistent with that obtained from immunofluorescence<sup>13</sup>.

**Electrophysiology.** Whole-cell patch clamping was performed with patch pipettes (4–9 M $\Omega$ ) containing Cs<sup>+</sup> internal solution (130 mM CsMeSO<sub>3</sub>, 10 mM Na-phosphocreatine, 4 mM MgCl<sub>2</sub>, 4 mM Na<sub>2</sub>-ATP, 0.4 mM Na<sub>2</sub>-GTP, 10 mM Cs-HEPES [pH 7.3])<sup>13</sup>. Excitatory postsynaptic current evoked by two-photon glutamate uncaging at a spine was measured through the patch pipette using a patch-clamp amplifier (Multiclamp 700B, Molecular Devices). Long-term potentiation was induced within 5 min of patching by pairing depolarization (0 mV) with two-photon glutamate uncaging at a spine (0.5 Hz, 4 ms, 30 pulses). Neurons showing more than 20% drift in the input or series resistances were not used for further analyses.

31. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296**, 913–916 (2002).
32. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnol.* **22**, 1567–1572 (2004).
33. Pédélecq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnol.* **24**, 79–88 (2005).
34. O'Brien, J. A. & Lummis, S. C. Biolistic transfection of neuronal cultures using a hand-held gene gun. *Nature Protocols* **1**, 977–981 (2006).
35. Yasuda, R. Imaging spatiotemporal dynamics of neuronal signaling using fluorescence resonance energy transfer and fluorescence lifetime imaging microscopy. *Curr. Opin. Neurobiol.* **16**, 551–561 (2006).
36. Murakoshi, H., Lee, S. J. & Yasuda, R. Highly sensitive and quantitative FRET-FLIM imaging in single dendritic spines using improved non-radiative YFP. *Brain Cell Biol.* **36**, 31–42 (2008).
37. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
38. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy* (Plenum, 2006).
39. Lee, S. J. & Yasuda, R. Spatiotemporal regulation of signaling in and out of dendritic spines. *CaMKII and Ras Open Neurosci. J.* **3**, 117–127 (2010).
40. Zhao, J., Wang, W. N., Tan, Y. C., Zheng, Y. & Wang, Z. X. Effect of Mg(2+) on the kinetics of guanine nucleotide binding and hydrolysis by Cdc42. *Biochem. Biophys. Res. Commun.* **297**, 653–658 (2002).

# Crystal structure of inhibitor of $\kappa$ B kinase $\beta$

Guozhou Xu<sup>1</sup>\*, Yu-Chih Lo<sup>1</sup>\*, Qiubai Li<sup>1</sup>, Gennaro Napolitano<sup>2</sup>, Xuefeng Wu<sup>2</sup>, Xuliang Jiang<sup>3</sup>, Michel Dreano<sup>4</sup>, Michael Karin<sup>2</sup> & Hao Wu<sup>1</sup>

**Inhibitor of  $\kappa$ B (I $\kappa$ B) kinase (IKK) phosphorylates I $\kappa$ B proteins, leading to their degradation and the liberation of nuclear factor  $\kappa$ B for gene transcription. Here we report the crystal structure of IKK $\beta$  in complex with an inhibitor, at a resolution of 3.6 Å. The structure reveals a trimodular architecture comprising the kinase domain, a ubiquitin-like domain (ULD) and an elongated,  $\alpha$ -helical scaffold/dimerization domain (SDD). Unexpectedly, the predicted leucine zipper and helix-loop-helix motifs do not form these structures but are part of the SDD. The ULD and SDD mediate a critical interaction with I $\kappa$ B $\alpha$  that restricts substrate specificity, and the ULD is also required for catalytic activity. The SDD mediates IKK $\beta$  dimerization, but dimerization per se is not important for maintaining IKK $\beta$  activity and instead is required for IKK $\beta$  activation. Other IKK family members, IKK $\alpha$ , TBK1 and IKK-i, may have a similar trimodular architecture and function.**

Nuclear factor  $\kappa$ B (NF- $\kappa$ B) transcription factors are master regulators of inflammatory, immune and apoptotic responses<sup>1,2</sup>. In the canonical pathway, NF- $\kappa$ B dimers are held in the cytoplasm through binding to I $\kappa$ B proteins, which mask their nuclear localization signals. When cells are stimulated by NF- $\kappa$ B inducers, I $\kappa$ B proteins are phosphorylated by the Ser/Thr-specific IKK, a modification that triggers their Lys-48-linked polyubiquitination and subsequent proteasomal degradation<sup>3</sup>. Freed NF- $\kappa$ B dimers enter the nucleus to regulate gene transcription<sup>2</sup>. In the non-canonical pathway, activated IKK phosphorylates the I $\kappa$ B-like domain in the NF- $\kappa$ B family member NF- $\kappa$ B2 (p49/p100)<sup>3</sup> (NFKB2 in humans). NF- $\kappa$ B signalling pathways are associated with a vast number of human diseases including inflammatory disorders and cancer, which renders IKK a potentially important therapeutic target<sup>4</sup> (see, for example, <http://www.nf-kb.org>).

IKK was originally purified from HeLa cells as a multiprotein complex that contains the kinase subunits IKK $\alpha$  (CHUK) and/or IKK $\beta$  (IKKBK), and the regulatory protein NEMO<sup>5–11</sup> (IKK $\gamma$ , or IKBK $\gamma$ ). IKK $\alpha$  and IKK $\beta$  both contain an amino-terminal kinase domain (KD), predicted leucine zipper (LZ) and helix-loop-helix (HLH) domains, and a carboxy-terminal NEMO-binding domain (Fig. 1a). IKK $\beta$  seems to have an additional ULD following the KD, which is not predicted in the corresponding region of IKK $\alpha$ <sup>2</sup>. The IKK-related kinases TBK1 (NAK) and IKK-i (IKK- $\epsilon$ , or IKBKE) seem to have a similar domain organization<sup>12</sup>. Whereas IKK $\beta$  mediates activation of the canonical NF- $\kappa$ B pathway in response to pro-inflammatory stimuli, IKK $\alpha$  has an indispensable role in non-canonical NF- $\kappa$ B signalling by phosphorylating NF- $\kappa$ B2 (p49/p100). Protein kinase assays suggest that IKK $\beta$  accounts for nearly all of the catalytic activity of the IKK holoenzyme towards I $\kappa$ B $\alpha$ <sup>3,13</sup> (NFKBIA).

The activation loop in both the IKK $\alpha$  and the IKK $\beta$  KD contains the mitogen-activated protein (MAP)-kinase kinase consensus motif SXXXS, where X is any amino acid<sup>6–8,10</sup> (Ser 177 and Ser 181 in human IKK $\beta$ ). Some MAP-kinase kinase family members, such as TGF- $\beta$ -activated kinase 1 (MAP3K7, previously TAK1) and NF- $\kappa$ B-inducing kinase (MAP3K14, or NIK), were shown to phosphorylate IKKs<sup>3,14,15</sup>. IKK $\alpha$  and IKK $\beta$  can also undergo autophosphorylation and activation as a result of overexpression or signal-dependent NEMO clustering<sup>10,16</sup>. Ala substitutions of the activation-loop Ser residues prevent IKK activation whereas the phosphomimetic, double

Glu mutation S177E/S181E (EE) of IKK $\beta$  renders it constitutively active<sup>7,13</sup>.

## Trimodular arrangement of IKK $\beta$

We determined the crystal structure of *Xenopus laevis* IKK $\beta$  (ikkbk) EE (residues 4–675; Fig. 1a) in complex with either inhibitor Cmpd1 or inhibitor Cmpd2 (Supplementary Fig. 1) at resolutions of 4.0 and 3.6 Å in the *I*<sub>4</sub>22 and *P*1 space groups, respectively (Supplementary Tables 1 and 2 and Supplementary Fig. 2). Eight IKK $\beta$  molecules in *P*1 and the single molecule in *I*<sub>4</sub>22 are highly similar to each other (Supplementary Fig. 3 and Supplementary Table 3) and show conserved dimerization (see below). Our structural description will use the first dimer (chains A and B) in *P*1. The sequences of human and *Xenopus* IKK $\beta$  (henceforth hIKK $\beta$  and xIKK $\beta$ , respectively) have 74% identity with no gaps within the region of the structure; residue numbers designated for xIKK $\beta$  are also true for hIKK $\beta$ .

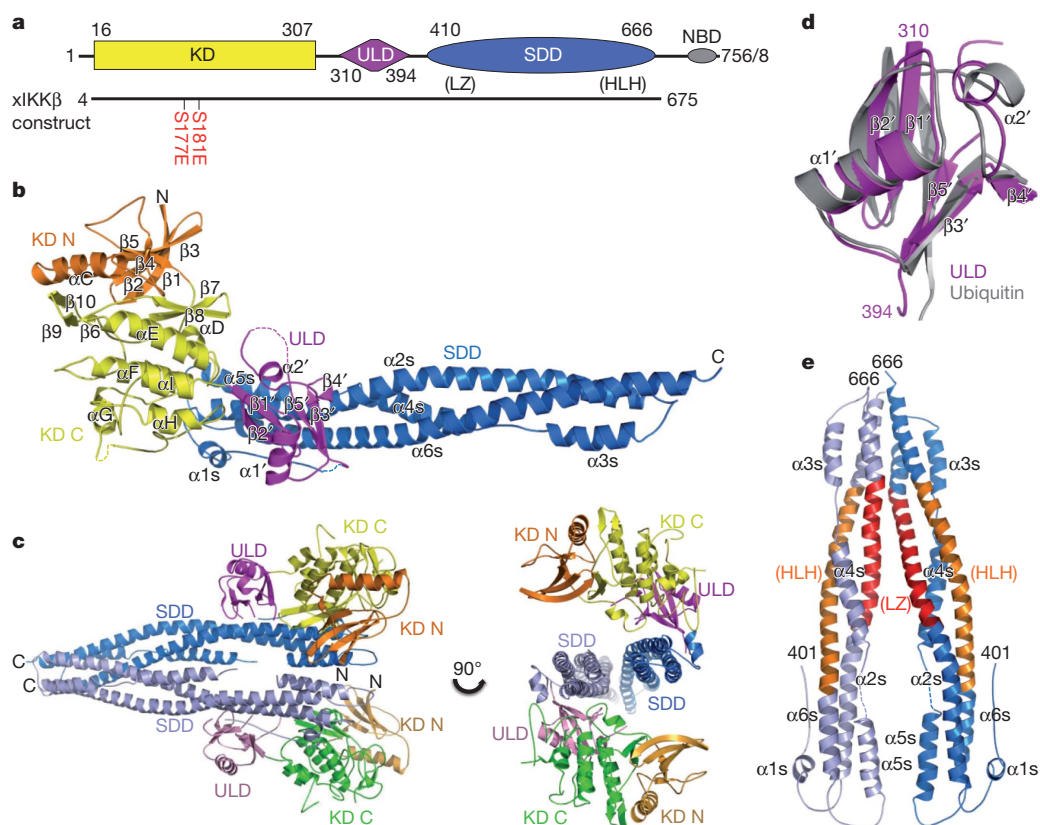
The IKK $\beta$  dimer structure resembles a pair of shears and has the overall dimensions of approximately 100 Å × 130 Å × 60 Å (Fig. 1b, c). It contains the KD (residues 16–307), the ULD (310–394) and the SDD (410–666) (Fig. 1a and Supplementary Fig. 4). The KD and the ULD form the 'handle' of the shears, and the SDD is the 'blade'. Both the KD and the ULD intimately associate with the SDD, suggesting that the KD does not function independently. Instead, the structure indicates that IKK $\beta$  is an integral trimodular unit.

The IKK $\beta$  KD in complex with either Cmpd1 or Cmpd2 has the typical bilobal kinase fold<sup>17</sup>. Although it has only 21.1% sequence identity with human ubiquitin, the ULD of IKK $\beta$  indeed has the ubiquitin fold (Fig. 1d). A major difference is the presence of an eight-residue insertion (373–380) that forms part of the loop connecting  $\beta$ -strands  $\beta$ 4' and  $\beta$ 5' in the ULD, which is mostly disordered. The hydrophobic surface patch of ubiquitin<sup>18</sup>, which is often recognized by ubiquitin-binding proteins, is conserved in the ULD, with residues Val 318, Leu 354 and Leu 389 equivalent to Leu 8, Ile 44 and Val 70 of ubiquitin (Supplementary Fig. 5).

The SDD comprises six  $\alpha$ -helices ( $\alpha$ 1s– $\alpha$ 6s), of which  $\alpha$ 2s and  $\alpha$ 6s contain 70 and 77 residues, respectively. They twist together with a stretch of three shorter helices between them to fold as an elongated structural ensemble. The C lobe of the KD sits on the N-terminal end

<sup>1</sup>Department of Biochemistry, Weill Cornell Medical College, New York, New York 10021, USA. <sup>2</sup>Department of Pharmacology, University of California at San Diego, La Jolla, California 92093, USA. <sup>3</sup>EMD Serono Research Institute, Billerica, Massachusetts 01821, USA. <sup>4</sup>Merck Serono, Geneva 1211, Switzerland.

\*These authors contributed equally to this work.

**Figure 1 | Structure of xIKKβ.**

**a**, Linear representation of IKKβ showing the boundaries for the KD, the ULD and the SDD. Sequences of hIKKβ and xIKKβ are of 756 and 758 residues, respectively, and differ only at the most C-terminal region. The crystallized xIKKβ construct is shown. The previously designated LZ and HLH regions are shown in parentheses. **b**, Ribbon diagram of an xIKKβ protomer in the P1 crystal form. The N and C termini, KD N lobe (orange) and C lobe (yellow), ULD (magenta) and SDD (blue) are labelled. Secondary structural elements are labelled, with those in the ULD indicated with a prime and those in the SDD indicated with an 's'. **c**, Ribbon diagram of an xIKKβ dimer. **d**, Superposition of the ULD (magenta) with ubiquitin (grey). **e**, Ribbon diagram of an SDD dimer, showing locations of the previously designated LZ (red) and HLH (orange) regions.

of the SDD, and the ULD binds close to the middle of the SDD. Unexpectedly, formerly designated LZ (458–485) and HLH domains<sup>7,8,10</sup> (605–644) do not exist as such in the structure and are both part of the SDD (Fig. 1e). Most of the hydrophobic residues in the predicted LZ point inwards and are not available for mediating dimerization as previously proposed.

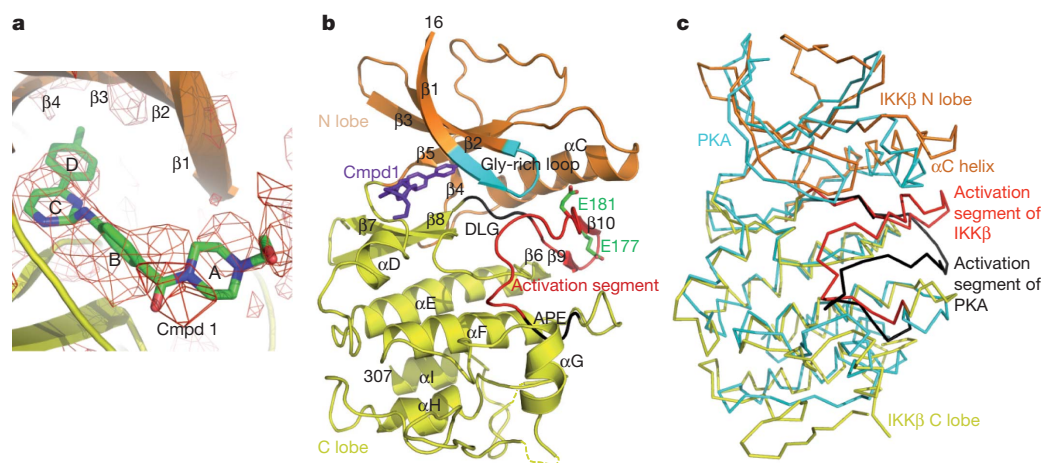
### Structure of inhibitor-bound IKKβ KD

The inhibitor binds to the IKKβ KD at the hinge loop connecting the N and C lobes, a region that recognizes the adenine in ATP<sup>19,20</sup> and is often used for inhibitor binding<sup>21–23</sup> (Fig. 2a and Supplementary Fig. 6). The KD conformation is incompatible with that of an active kinase<sup>17,24,25</sup> (Fig. 2b, c). The activation segment, which begins from Asp 166–Leu 167–Gly 168 as the conserved DFG triad and extends to Ala 190–Pro 191–Glu 192 in the conserved APE motif, is fully ordered, including

phosphomimetic residues Glu 177 and Glu 181 (Fig. 2b). However, the C-terminal anchor of the activation segment, including the APE motif, is completely out of place, with the result that essential interactions with the catalytic centre are lost (Fig. 2c). The gross conformation of the N-terminal anchor of the activation segment is preserved, with β9 paired with the β6 strand that precedes the catalytic loop. Part of the activation loop (175–177) contains an additional β-strand (β10), which sits between the lobes to form a three-stranded β-sheet with β9 and β6. The αC helix is tilted up and outwards (Fig. 2c) to make room for β10, which also disrupts its productive interactions with the DFG motif in active kinase structures.

### Interactions among the KD, ULD and SDD

The KD, ULD and SDD interact mutually, with the ULD–SDD interaction being the most extensive, followed by the KD–SDD and KD–ULD

**Figure 2 | Inhibitor-bound xIKKβ**

**kinase domain.** **a**,  $F_o - F_c$  electron density map for Cmpd1 in the I4/22 structure, contoured at  $2.0\sigma$ . Carbon, nitrogen and oxygen atoms are shown in green, blue and red, respectively. The four ring structures in Cmpd1 are labelled A, B, C and D, respectively. **b**, Structure of the xIKKβ KD. Gly-rich loop, cyan; activation segment, red (except that the DLG and APE sequences are in black); Cmpd1, purple. Side chains of phosphomimetic residues Glu 177 and Glu 181 are shown. **c**, Superposition of xIKKβ (orange and yellow) and protein kinase A (PKA, cyan; Protein Data Bank ID, 1ATP). The activation segments of xIKKβ and PKA are shown in red and black, respectively.



interactions. The ULD binds close to the middle of the SDD, contacting helices  $\alpha 2s$  and  $\alpha 6s$  (Fig. 3a). The interaction buries  $\sim 700 \text{ \AA}^2$  of surface area per partner. There are substantial hydrophobic contributions, supplemented by electrostatic interactions. Residues Met 315, Met 317, Leu 354, Ile 387, Leu 389 and Phe 390 on one side of the ULD pack against Leu 612, Tyr 609, Leu 447 and the main chain of  $\alpha 2s$  of the SDD to form the central hydrophobic core of the interface. This hydrophobic patch of the ULD is immediately adjacent to and overlaps the conserved hydrophobic patch in the ubiquitin fold. Electrostatic interactions are observed between Glu 352 and Lys 619 and between Lys 394 and Asp 610. Additional interfacial residues include Ser 319 and Ser 357 of the ULD and Thr 453, Gln 449 and Val 616 of the SDD.

Consistent with an important role of the ULD in interacting with the SDD, mutations in residues that are not at the interface, P347A, L361A and Q351A, had minimal effects on NF- $\kappa$ B activation<sup>26</sup>. In contrast, mutation in a residue within a ULD surface loop that contacts the SDD, G358A (Fig. 3a), affected IKK $\beta$ -induced NF- $\kappa$ B activity<sup>26</sup>. It was also shown that Leu 353 is required for IKK $\beta$  activity<sup>26</sup>; however, Leu 353 is buried in the ULD core and the L353A mutation may have disrupted IKK $\beta$  structural integrity. Double substitutions of hIKK $\alpha$  and hIKK $\beta$ , which are equivalent to I608R/Y609P of the SDD of  $\alpha$ IKK $\beta$ , did not affect their interaction with wild-type IKK $\beta$  but negatively impacted kinase activity<sup>10</sup>; Ile 608 is buried in the SDD core and Tyr 609 directly interacts with the ULD (Fig. 3a).

Like the ULD, the KD also makes contact with the N terminal end of the SDD (Fig. 3b), burying a surface area of  $\sim 650 \text{ \AA}^2$  of each interface. The binding forces are mainly van der Waals in nature. Limited hydrophobic interactions are observed between residues Ala 252 and Val 253 of the KD and Tyr 423 of the SDD, and between Phe 111 of the KD and the hydrophobic portions of Arg 572, Arg 575 and Glu 576 of  $\alpha 5s$  of the SDD. The KD is linked to the ULD through a two-amino-acid linker between  $\alpha I$  of the KD and  $\beta 1'$  of the ULD (Fig. 3c), burying only  $\sim 350 \text{ \AA}^2$  of surface area. Side-chain contacts between Leu 311 of the ULD and Ile 268 of the KD are observed, and together form the small hydrophobic patch at the KD–ULD junction

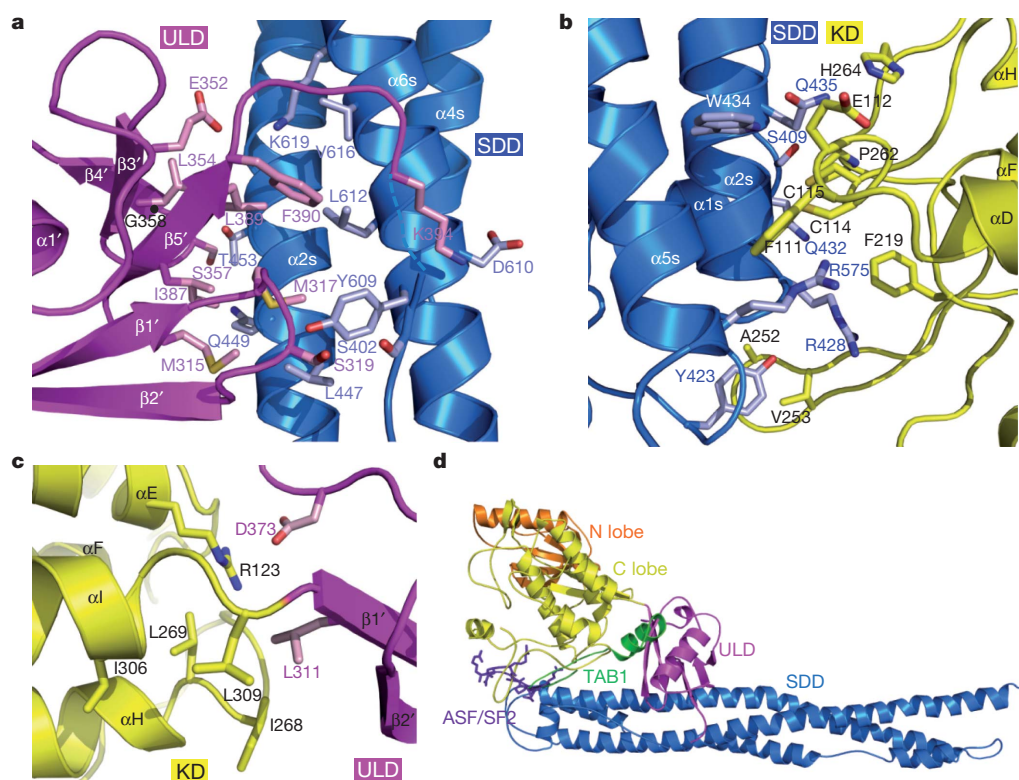
consisting also of Leu 269 and Ile 306 of the KD and Leu 309 of the linker. Together with the ionic interaction between Asp 373 of the ULD and Arg 123 of  $\alpha E$  of the KD, these interactions may confer rigidity to the KD–ULD junction despite the small buried surface area.

Structural comparison with other kinase structures revealed a similarity between the locations of the SDD and ULD and those of several known docking sites for substrates and regulatory proteins<sup>27</sup>. In the crystal structure of the Ser/Thr kinase SRPK1 in complex with a docking motif in its substrate, ASF/SF2<sup>27</sup> (SRSF1), the peptide motif interacts with the kinase at the location of the SDD (Fig. 3d). In the structure of the TAK1 KD fused with the C-terminal region of its binding protein, TAB1<sup>28</sup>, TAB1 interacts with the C lobe of the kinase at a position analogous to both the SDD and the ULD, presumably to activate the kinase (Fig. 3d).

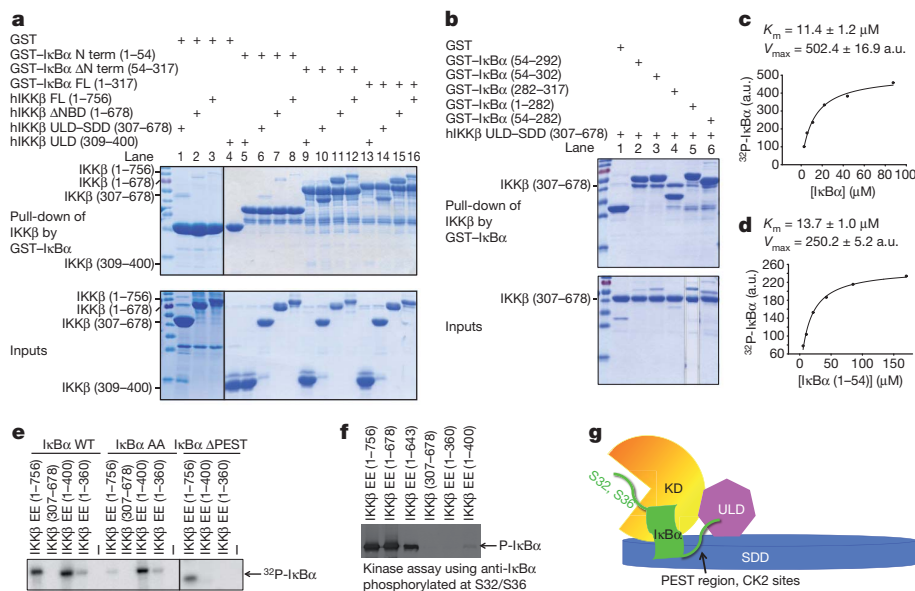
### ULD–SDD binds the I $\kappa$ B $\alpha$ C-terminal region

Previous studies have suggested that the ULD in TBK1 and IKK-i is involved in substrate recognition because its deletion impaired activity of the respective kinases and a glutathione *S*-transferase (GST)–ULD fusion protein interacted with the specific substrate, IRF3 or IRF7<sup>29</sup>. Because ULD deletion in IKK $\beta$  also abolished its activity<sup>26</sup>, we proposed that the ULD may recognize its specific substrate, I $\kappa$ B $\alpha$ . However, we were surprised to find that GST–I $\kappa$ B $\alpha$  (1–317 and 54–317) pulled down only a minute amount of the ULD (Fig. 4a, lanes 9 and 13) and GST alone did not pull down any (Fig. 4a, lane 4), suggesting that the interaction of I $\kappa$ B $\alpha$  with the ULD is specific, but very weak. In contrast, GST–I $\kappa$ B $\alpha$  robustly pulled down full-length IKK $\beta$  or IKK $\beta$  lacking C-terminal NEMO-binding domain (Fig. 4a, lanes 15 and 16).

I $\kappa$ B $\alpha$  has an N-terminal region (1–54) that contains cognate phosphoacceptor sites at Ser 32 and S 36, followed by a C-terminal region (55–317) that contains multiple ankyrin repeats and the PEST region<sup>30,31</sup>. Strikingly, the N-terminal region of I $\kappa$ B $\alpha$  did not pull down any IKK $\beta$  constructs (Fig. 4a, lanes 5–8), and the C-terminal region of I $\kappa$ B $\alpha$  interacted robustly with full-length IKK $\beta$  as well as its ULD–SDD region (Fig. 4a, lanes 9–12), and very weakly with the ULD alone (Fig. 4a, lane 9). Further mapping on I $\kappa$ B $\alpha$  showed that both



**Figure 3 | Interactions among the KD, ULD and SDD.** **a**, Interaction between the ULD (magenta) and SDD (blue). Important interfacial side chains are shown with nitrogen atoms in blue, oxygen atoms in red, sulphur atoms in orange and carbon atoms in either pink (ULD) or light blue (SDD). Gly 358 is marked with a black ball on the main chain. **b**, Interaction between the KD (yellow) and SDD (blue). **c**, Interaction between the KD (yellow) and ULD (magenta). **d**, Locations of the TAB1 peptide (green ribbon; PDB ID, 2EVA) and the ASF/SF2 peptide (purple stick model; PDB ID, 1WBP) relative to the IKK $\beta$  structure after superposition of the corresponding kinase domains.



**Figure 4** | ULD-SDD restricts IKKβ specificity and ULD is required for catalytic activity.

**a, b**, Pull-down of hIKKβ constructs using GST-IκBα constructs, showing the reciprocal interaction between ULD-SDD of IKKβ and the C-terminal region of IκBα containing ankyrin repeats and the PEST region. FL, full length. **c, d**, Measurement of  $K_m$  and the relative  $V_{max}$  of IKKβ against full-length IκBα (**c**) and the N-terminal region of IκBα (1–54) (**d**). a.u., arbitrary units. **e**, Kinase assay of purified hIKKβ proteins against IκBα, its S32A/S36A mutant (AA) and its PEST-deletion construct (ΔPEST, 1–282) using [ $\gamma$ - $^{32}P$ ]ATP. WT, wild type. **f**, Kinase assay of purified hIKKβ proteins using antibody against IκBα phosphorylated at Ser 32 and Ser 36. **g**, Interaction between the SDD of IKKβ and the C-terminal region of IκBα may position the N-terminal cognate phosphorylation sites of IκBα at the active site of IKKβ. CK2, casein kinase 2.

ankyrin repeats (1–282 or 54–282) and the PEST region (282–317) interacted with IKKβ ULD-SDD (Fig. 4b, lanes 4–6). The extent of pull-down suggests that the PEST region contributes more to IKKβ interaction than do the ankyrin repeats. Despite trying multiple constructs, we could not obtain soluble protein for IKKβ SDD to test its interaction with IκBα. However, the pull-down data suggest a dominant role for SDD in IκBα interaction. In any case, it is clear that the mutual interaction between IKKβ and IκBα is mediated by their ULD-SDD and C-terminal domains, respectively.

### ULD-SDD limits specificity and the ULD aids catalysis

The specific interaction between ULD-SDD of IKKβ and IκBα suggests that the Michaelis constant,  $K_m$ , of phosphorylation by IKKβ might be lower for full-length IκBα than for its N-terminal region (1–54) alone. We performed kinase kinetics analysis of IKKβ EE against the two different substrates. Unexpectedly, the measured apparent  $K_m$  values were 11.4 and 13.7  $\mu M$  for full-length IκBα and the N-terminal region alone, respectively (Fig. 4c, d), suggesting that binding at the SDD, an exosite, does not alter the  $K_m$  of the reaction drastically. This could be due to opposing effects of the SDD-IκBα interaction, which increases substrate interaction but slows down product dissociation. The relative maximum velocity,  $V_{max}$  values were 502.4 and 250.2, respectively, suggesting that the SDD-IκBα interaction moderately enhances catalysis.

Casein kinase 2 phosphorylates the C-terminal PEST region of IκBα around residues 283–299<sup>30,31</sup>. To determine whether the SDD-IκBα interaction restricts substrate specificity, we compared the kinase activity of purified IKKβ EE proteins against either IκBα or its S32A/S36A mutant (AA) using [ $\gamma$ - $^{32}P$ ]ATP (Fig. 4e). Although the KD-ULD (1–400) construct gave rise to a small amount of protein, it produced robust phosphorylation of wild-type IκBα, comparable to full-length IKKβ (1–756), suggesting that it is catalytically competent. Remarkably, KD-ULD effectively phosphorylated the AA mutant, which, in contrast, was a poor substrate for full-length IKKβ. The C-terminal PEST region seemed to harbour the major sites of phosphorylation by KD-ULD because a construct lacking PEST (1–282) was not significantly phosphorylated by KD-ULD but was phosphorylated by full-length IKKβ (Fig. 4e). In addition, when IκBα phosphorylation was detected by antibody against IκBα phosphorylated at Ser 32/Ser 36, only very weak phosphorylation was seen for the KD-ULD construct in comparison with full-length IKKβ (Fig. 4f). These experiments suggest that although the KD-ULD construct is active, it possesses an altered substrate specificity causing preferential

phosphorylation of the C-terminal PEST sequence, consistent with a previous observation<sup>32</sup>. Hence, ULD-SDD seems to position IκBα specifically such that only its N-terminal region is accessible to the IKKβ catalytic pocket (Fig. 4g).

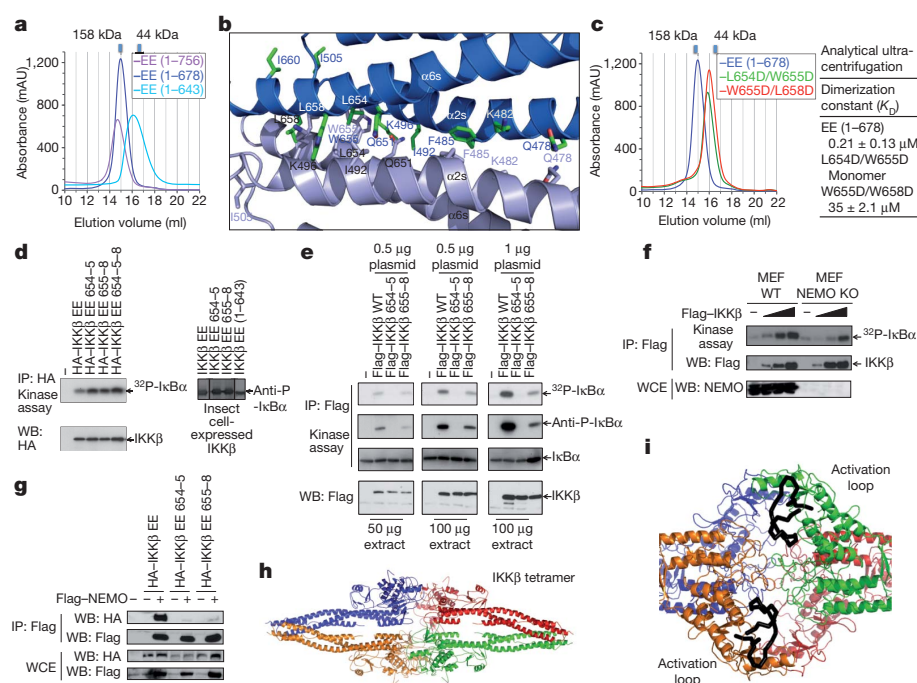
We expressed three kinase domain constructs, 1–301, 1–310 and 1–360, in insect cells and obtained only small amount of protein for the 1–360 construct. Kinase assay showed that IKKβ EE (1–360) had very low residual activity against IκBα or its AA mutant (Fig. 4e), suggesting that KD activity is compromised in the absence of the ULD. Without an isolated KD structure, we cannot deduce the molecular mechanism by which the ULD acts. However, in analogy to TAK1 activation by TAB1<sup>28</sup> (Fig. 3d), it may be speculated that this KD-ULD interaction allosterically potentiates kinase activity. Further kinase assay using antibody against IκBα phosphorylated at Ser 32/Ser 36 showed no detectable activity of the KD alone (Fig. 4f), confirming that the low residual activity may also be directed towards the C-terminal PEST region, similar to KD-ULD. Therefore, whereas ULD-SDD is critical for IKKβ specificity, ULD is required for its catalytic activity.

### SDD mediates IKKβ dimerization

Full-length hIKKβ (1–756) and its KD-ULD-SDD region (1–678) are dimers in solution as determined by gel-filtration chromatography (Fig. 5a). In the P1 and I4122 crystals, two conserved dimerization interfaces exist, one mediated by the SDD and the other mediated by the KD. Because hIKKβ (1–643) that truncates into the SDD is a monomer in solution (Fig. 5a), SDD-mediated dimerization (Fig. 5b) is probably of greater importance than KD-mediated dimerization.

SDD in an IKKβ dimer does not form extensive interactions along the entire length dimension of the helical bundle. Rather, interactions are mostly localized at the end of the bundle (Fig. 5b), distal from the KD and ULD and burying  $\sim 1,000 \text{ \AA}^2$  of surface area of each monomer. The interaction is mostly hydrophobic. Residues that contribute significantly to dimerization include Gln 478, Lys 482, Phe 485, Ile 492, Lys 496, Ile 505, Gln 651, Leu 654, Trp 655, Leu 658 and Ile 660, with residues Leu 654, Trp 655 and Leu 658 from helix  $\alpha 6$  burying the most surface area (Fig. 5b). This dimerization interface is entirely unexpected as it was predicted earlier that the LZ mediates IKKβ dimerization. The structure now explains that failed dimerization of the LZ-defective L462S/L469S mutant of IKKα<sup>10</sup> is due to the structural role of Leu 469, whose equivalent residue of IKKβ, Met 472, is buried in the SDD core. Superposition of four IKKβ dimers in P1





**Figure 5 | Dimerization is critical for IKK $\beta$  activation but not for its activity.** **a**, Gel filtration profiles of various hIKK $\beta$  constructs showing that those containing the KD–ULD–SDD region (1–756, 1–678) are dimeric and that a further truncated construct (1–643) is monomeric. mAU, milliabsorbance unit. **b**, The dimerization interface of xIKK $\beta$ . **c**, Structure-based mutations disrupt hIKK $\beta$  dimerization as shown by gel filtration and analytical ultracentrifugation. **d**, Kinase activity of HEK293T-cell-transfected or insect-cell-purified hIKK $\beta$  EE and dimerization-defective mutants L654D/W655D (654–655), W655D/L658D (655–658) and L654D/W655D/L658D (654–655–658). **e**, Autoactivation of HEK293T-cell-transfected wild-type hIKK $\beta$  and its dimerization-defective mutants. **f**, Transfection of hIKK $\beta$  into wild-type and NEMO $^{-/-}$  mouse embryonic fibroblasts (MEFs), showing reduced IKK $\beta$  activation in the absence of NEMO. **g**, Dimerization mutants of IKK $\beta$  showing reduced interaction with NEMO. **h**, Tetramer of xIKK $\beta$  in the P1 structure. **i**, Close-up view of the tetramer interface, showing that the activation loops of neighbouring protomers (black) face each other.

and the single IKK $\beta$  dimer in *I4,22* shows that IKK $\beta$  dimers are conserved (Supplementary Fig. 7). In *I4,22*, the distal part of the SDD is not visible, owing to a lack of crystal packing in this region and dynamic disorder, not degradation (Supplementary Fig. 8).

To confirm the observed interface in IKK $\beta$  dimerization, we performed structure-based mutagenesis on residues that bury the most surface area, Leu 654, Trp 655 and Leu 658. These residues are not within the predicted LZ region. Two purified double mutants, L654D/W655D and W655D/L658D, failed to dimerize, as shown by the considerable shift in gel filtration positions (Fig. 5c). Furthermore, equilibrium analytical ultracentrifugation experiments showed that wild-type IKK $\beta$  is indeed a dimer whereas one of the IKK $\beta$  mutants is a monomer and the other has a 167-fold weaker dimerization affinity (Fig. 5c).

### Dimerization in IKK $\beta$ activation but not activity

The geometry of the IKK $\beta$  dimer with KDs facing away from each other suggests that each IKK $\beta$  protomer is independent in its kinase activity. To confirm this, we transfected HEK293T cells with hIKK $\beta$  EE mutants, L654D/W655D (654–655), W655D/L658D (655–658) and L654D/W655D/L658D (654–655–658), and found that all mutants had robust kinase activity (Fig. 5d). Dimerization mutants expressed in insect cell and purified showed the same results. In addition, a purified IKK $\beta$  construct with truncation into the SDD (1–643) that is monomeric in solution (Fig. 5a) was active in phosphorylating I $\kappa$ B $\alpha$  (Fig. 5d).

Previous studies have shown that IKK $\beta$  can undergo *trans*-autophosphorylation and activation on transfection<sup>13,16</sup>. This autophosphorylation and phosphorylation by TAK1 probably both contribute to IKK $\beta$  activation on cell stimulation. To determine whether the observed dimerization interface is critical for this activation, we tested dimerization mutations in IKK $\beta$  activation on overexpression in HEK293T cells (Fig. 5e). Whereas wild-type IKK $\beta$  was robustly activated, the L654D/W655D and W655D/L658D mutants completely and partly lost this activation, respectively, in a manner that correlates with the degree of impairment in dimerization (Fig. 5c, e). Overexpression of IKK $\beta$  in NEMO-deficient mouse embryonic fibroblasts resulted in its activation, but to a lesser extent than in wild-type mouse embryonic fibroblasts (Fig. 5f). We found that IKK $\beta$  dimerization mutants failed to interact with NEMO efficiently (Fig. 5g), perhaps owing to reduced

avidity for oligomeric NEMO. Therefore, although IKK $\beta$  kinase activity does not depend on dimerization once its activation loop is phosphorylated, IKK $\beta$  activation requires dimerization and is probably enhanced by interaction with NEMO.

### Discussion

The IKK $\beta$  structure presented here predicts that IKK $\alpha$ , TBK1 and IKK-i all have an overall structural organization that comprises KD, ULD and SDD. Although a ULD was not predicted in IKK $\alpha$ , conservation of buried residues in this region and of the ULD–SDD interface suggests that IKK $\alpha$  also has this domain (Supplementary Fig. 4). The ULD and SDD probably have the same structural role but may have evolved additional, differential functions in the individual kinases. SDD-mediated dimerization may also be conserved. In particular, residues at the observed IKK $\beta$  dimerization interface are highly conserved in IKK $\alpha$  (Supplementary Fig. 4), explaining how IKK $\alpha$  and IKK $\beta$  can form both homo- and heterodimers<sup>8,13</sup>. Given our structure of IKK $\beta$  and the previously determined structures of NEMO fragments<sup>33–37</sup>, we may speculate that the full IKK complex is also a dimer, or a dimer of dimers with a molecular mass of around 270 or 540 kD. The apparent 700–900-kD molecular mass of the IKK holoenzyme on gel filtration<sup>5–10</sup> may be due to the elongated nature of NEMO and the complex (Supplementary Fig. 9).

Because the conserved IKK $\beta$  dimer structure does not place the KDs close to each other for *trans*-autophosphorylation, we wondered whether higher-order oligomerization, which is probably enhanced by NEMO and its ability to bind ubiquitin<sup>33,34</sup>, is responsible for this autoactivation. In both *P1* and *I4,22*, IKK $\beta$  exists as dimers of dimers (Fig. 5h and Supplementary Fig. 10). In particular, active sites of two neighbouring protomers in the tetramer face each other such that an activation loop from one protomer might reach into the active site of the other (Fig. 5i), which may mimic an autophosphorylation state.

The ULD–SDD region of IKK $\beta$  directly interacts with the C-terminal portion of I $\kappa$ B $\alpha$ . This interaction may serve several purposes. First, it probably orients I $\kappa$ B $\alpha$  such that its N-terminal cognate phosphorylation sites are presented to the KD active site (Fig. 4g). Without this interaction, IKK $\beta$  preferentially phosphorylates the C-terminal PEST motif of I $\kappa$ B $\alpha$ . Second, the interaction seems to enhance IKK $\beta$  activity. Third, phosphorylation at the PEST by casein kinase 2 or other kinases may regulate I $\kappa$ B $\alpha$  interaction with IKK $\beta$



and hence affect phosphorylation at the cognate sites. Fourth, the interaction may allow concerted phosphorylation at both Ser 32 and Ser 36 of I $\kappa$ B $\alpha$  without intervening dissociation. In MAP-kinase cascades that involve dual phosphorylations, specific docking interactions occur between the kinases and their substrates<sup>27,38,39</sup>. The  $\beta$ -catenin protein in the Wnt signalling pathway contains the same dual-phosphorylated destruction box motif as I $\kappa$ B $\alpha$  (ref. 40). Consistent with this,  $\beta$ -catenin is brought to the responsible kinase, GSK-3, by means of the adaptor protein axin, allowing both specificity and concerted phosphorylation<sup>41,42</sup>. Therefore, in a general statement that structure serves function, the IKK $\beta$  structure seems to fit its function as the I $\kappa$ B $\alpha$  kinase perfectly, being poised to turn on the important NF- $\kappa$ B pathway specifically, efficiently and concertedly in response to cellular physiology.

## METHODS SUMMARY

*Xenopus laevis* IKK $\beta$  was expressed in insect cells and purified to homogeneity using nickel affinity chromatography, ion exchange and gel filtration chromatography. We crystallized the P1 and I4,22 forms at 4 °C in polyethylene glycol 6000 and potassium/sodium phosphate, respectively. The structure was determined by multiwavelength anomalous diffraction of the selenomethionyl protein.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 September 2010; accepted 17 January 2011.

Published online 20 March 2011.

- Hayden, M. S. & Ghosh, S. Shared principles in NF- $\kappa$ B signaling. *Cell* **132**, 344–362 (2008).
- Vallabhupuram, S. & Karin, M. Regulation and function of NF- $\kappa$ B transcription factors in the immune system. *Annu. Rev. Immunol.* **27**, 693–733 (2009).
- Scheidereit, C. I $\kappa$ B kinase complexes: gateways to NF- $\kappa$ B activation and transcription. *Oncogene* **25**, 6685–6705 (2006).
- Karin, M. Nuclear factor- $\kappa$ B in cancer development and progression. *Nature* **441**, 431–436 (2006).
- Chen, Z. J., Parent, L. & Maniatis, T. Site-specific phosphorylation of I $\kappa$ B $\alpha$  by a novel ubiquitination-dependent protein kinase activity. *Cell* **84**, 853–862 (1996).
- DiDonato, J. A. *et al.* A cytokine-responsive I $\kappa$ B kinase that activates the transcription factor NF- $\kappa$ B. *Nature* **388**, 548–554 (1997).
- Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated I $\kappa$ B kinases essential for NF- $\kappa$ B activation. *Science* **278**, 860–866 (1997).
- Woronicz, J. D. *et al.* I $\kappa$ B kinase- $\beta$ : NF- $\kappa$ B activation and complex formation with I $\kappa$ B kinase- $\alpha$  and NIK. *Science* **278**, 866–870 (1997).
- Yamaoka, S. *et al.* Complementation cloning of NEMO, a component of the I $\kappa$ B kinase complex essential for NF- $\kappa$ B activation. *Cell* **93**, 1231–1240 (1998).
- Zandi, E. *et al.* The I $\kappa$ B kinase complex (IKK) contains two kinase subunits, IKK $\alpha$  and IKK $\beta$ , necessary for I $\kappa$ B phosphorylation and NF- $\kappa$ B activation. *Cell* **91**, 243–252 (1997).
- Rothwarf, D. M., Zandi, E., Natoli, G. & Karin, M. IKK- $\gamma$  is an essential regulatory subunit of the I $\kappa$ B kinase complex. *Nature* **395**, 297–300 (1998).
- Hacker, H. & Karin, M. Regulation and function of IKK and IKK-related kinases. *Sci. STKE* **2006**, re13 (2006).
- Zandi, E., Chen, Y. & Karin, M. Direct phosphorylation of I $\kappa$ B by IKK $\alpha$  and IKK $\beta$ : discrimination between free and NF- $\kappa$ B-bound substrate. *Science* **281**, 1360–1363 (1998).
- Sato, S. *et al.* Essential function for the kinase TAK1 in innate and adaptive immune responses. *Nature Immunol.* **6**, 1087–1095 (2005).
- Liu, H. H., Xie, M., Schneider, M. D. & Chen, Z. J. Essential role of TAK1 in thymocyte development and activation. *Proc. Natl Acad. Sci. USA* **103**, 11677–11682 (2006).
- Tang, E. D. *et al.* Roles for homotypic interactions and transautophosphorylation in I $\kappa$ B kinase (IKK $\beta$ ) activation. *J. Biol. Chem.* **278**, 38566–38570 (2003); erratum **278**, 49661 (2003).
- Knighton, D. R. *et al.* Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–414 (1991).
- Dikic, I., Wakatsuki, S. & Walters, K. J. Ubiquitin-binding domains — from structures to functions. *Nature Rev. Mol. Cell Biol.* **10**, 659–671 (2009).
- Zheng, J. *et al.* 2.2 Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr. D* **49**, 362–365 (1993).
- Bossemeyer, D. *et al.* Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn<sup>2+</sup> adenylate
- imodidiphosphate and inhibitor peptide PKI(5–24). *EMBO J.* **12**, 849–859 (1993).
- Xu, R. M., Carmel, G., Kuret, J. & Cheng, X. Structural basis for selectivity of the isoquinoline sulfonamide family of protein kinase inhibitors. *Proc. Natl Acad. Sci. USA* **93**, 6308–6313 (1996).
- Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602–609 (1997).
- Noble, M. E., Endicott, J. A. & Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **303**, 1800–1805 (2004).
- Nolen, B., Taylor, S. & Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15**, 661–675 (2004).
- Jeffrey, P. D. *et al.* Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex. *Nature* **376**, 313–320 (1995).
- May, M. J. *et al.* A novel ubiquitin-like domain in I $\kappa$ B kinase  $\beta$  is required for functional activity of the kinase. *J. Biol. Chem.* **279**, 45528–45539 (2004).
- Goldsmith, E. J. *et al.* Substrate and docking interactions in serine/threonine protein kinases. *Chem. Rev.* **107**, 5065–5081 (2007).
- Brown, K. *et al.* Structural basis for the interaction of TAK1 kinase with its activating protein TAB1. *J. Mol. Biol.* **354**, 1013–1020 (2005).
- Ikeda, F. *et al.* Involvement of the ubiquitin-like domain of TBK1/IKK-i kinases in regulation of IFN-inducible genes. *EMBO J.* **26**, 3451–3462 (2007).
- Kato, T. Jr, Delhase, M., Hoffmann, A. & Karin, M. CK2 is a C-terminal I $\kappa$ B kinase responsible for NF- $\kappa$ B activation during the UV response. *Mol. Cell* **12**, 829–839 (2003).
- Barroga, C. F., Stevenson, J. K., Schwarz, E. M. & Verma, I. M. Constitutive phosphorylation of I $\kappa$ B $\alpha$  by casein kinase II. *Proc. Natl Acad. Sci. USA* **92**, 7637–7641 (1995).
- Shaul, J. D., Farina, A. & Huxford, T. The human IKK $\beta$  subunit kinase domain displays CK2-like phosphorylation specificity. *Biochem. Biophys. Res. Commun.* **374**, 592–597 (2008).
- Lo, Y. C. *et al.* Structural basis for recognition of diubiquitins by NEMO. *Mol. Cell* **33**, 602–615 (2009).
- Rahighi, S. *et al.* Specific recognition of linear ubiquitin chains by NEMO is important for NF- $\kappa$ B activation. *Cell* **136**, 1098–1109 (2009).
- Rushe, M. *et al.* Structure of a NEMO/IKK-associating domain reveals architecture of the interaction site. *Structure* **16**, 798–808 (2008).
- Bagnéris, C. *et al.* Crystal structure of a vFlip-IKK $\gamma$  complex: insights into viral activation of the IKK signalosome. *Mol. Cell* **30**, 620–631 (2008).
- Cordier, F. *et al.* Solution structure of NEMO zinc finger and impact of an anhidrotic ectodermal dysplasia with immunodeficiency-related point mutation. *J. Mol. Biol.* **377**, 1419–1432 (2008).
- Remenyi, A., Good, M. C. & Lim, W. A. Docking interactions in protein kinase and phosphatase networks. *Curr. Opin. Struct. Biol.* **16**, 676–685 (2006).
- Kallunki, T., Deng, T., Hibi, M. & Karin, M. c-Jun can recruit JNK to phosphorylate dimerization partners via specific docking interactions. *Cell* **87**, 929–939 (1996).
- Wu, G. *et al.* Structure of a  $\beta$ -TrCP1-Skp1- $\beta$ -catenin complex: destruction motif binding and lysine specificity of the SCF( $\beta$ -TrCP1) ubiquitin ligase. *Mol. Cell* **11**, 1445–1456 (2003).
- Ikeda, S. *et al.* Axin, a negative regulator of the Wnt signaling pathway, forms a complex with GSK-3 $\beta$  and  $\beta$ -catenin and promotes GSK-3 $\beta$ -dependent phosphorylation of  $\beta$ -catenin. *EMBO J.* **17**, 1371–1384 (1998).
- Hart, M. J. *et al.* Downregulation of  $\beta$ -catenin by human Axin and its association with the APC tumor suppressor,  $\beta$ -catenin and GSK3 $\beta$ . *Curr. Biol.* **8**, 573–581 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank K. Rajashankar and N. Sukumar for data collection at the NE-CAT of APS, B. Schwer for help with the kinase assay, P. Gaillard for help with the chemistry and G. Ahlsen, L. Shapiro and B. Honig for the ultracentrifugation experiments. This work was supported by the National Institutes of Health (H.W. and M.K.), the American Heart Association (G.X. and Y.-C.L.) and the Cancer Research Institute (Y.-C.L.). M.K. is an American Cancer Society Research Professor.

**Author Contributions** G.X. cloned, expressed, purified, crystallized and determined the crystal structure of xIKK $\beta$  and performed experiments to determine  $K_m$ . Y.-C.L. cloned, expressed, purified and crystallized hIKK $\beta$  and performed pull-down experiments and kinase assays using phospho-I $\kappa$ B $\alpha$  antibody. Q.L. expressed the hIKK $\beta$  mutants in insect cells. G.N. and X.W. performed transfection, immunoprecipitation and kinase assays and M.K. supervised these experiments. H.W. supervised the project. G.X. and H.W. made the figures and wrote the manuscript.

**Author Information** Atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 3QA8 and 3QAD. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to H.W. ([haowu@med.cornell.edu](mailto:haowu@med.cornell.edu)).

## METHODS

**Protein expression and purification.** To elucidate the molecular basis of IKK $\beta$  function, we expressed IKK $\beta$  from a number of species using baculovirus-mediated insect cell expression. The xIKK $\beta$  sequence in the NCBI database starts at a Met residue that is equivalent to Met 17 of both the hIKK $\beta$  and the mouse IKK $\beta$  (mIKK $\beta$ ) sequence. Translation of the DNA sequence preceding the ATG codon of Met 17 revealed sequences that are almost identical to residues 9–16 of hIKK $\beta$  and mIKK $\beta$ . These were taken as part of the xIKK $\beta$  sequence and residues 1–8 were taken from the corresponding mIKK $\beta$  sequence. This reconstructed xIKK $\beta$  sequence has the same residue numbering as the hIKK $\beta$  sequence until after the SDD.

Various constructs of IKK $\beta$  wild type and the phosphomimetic S177E/S181E mutant were designed with an N-terminal polyhistidine tag and a tobacco etch virus protease-cutting site between the tag and the protein. Recombinant IKK $\beta$  baculoviruses were made in DH10BAC cells, amplified and used to infect Hi5 insect cells in serum-free media (Invitrogen). The cells were cultured in suspension and harvested 48 h after infection. The recombinant proteins were purified by nickel affinity chromatography, anion exchange and gel filtration chromatography. For crystallization, the polyhistidine tag was cleaved by the tobacco etch virus protease during protein purification.

All IkB $\alpha$  proteins were expressed in *Escherichia coli* using pET28a, pGEX4T3 and pET-SUMO vectors and purified by their respective affinity tags. For kinase assays, the SUMO tag was cleaved from IkB $\alpha$  proteins. His-ULD and His-ULD-SDD of hIKK $\beta$  were also expressed in *E. coli* using the pET28a vector.

**Crystallization and data collection.** Unlike many protein kinases, the IKK $\beta$  KD cannot be recombinantly expressed as a well-behaved biochemical entity for structural studies. In addition, after mapping a compact region by limited proteolysis, IKK $\beta$  was still refractory to crystallization, both alone and in the presence of various ATP analogues. To overcome this obstacle, we used several IKK $\beta$  inhibitors, including Cmpd1 and Cmpd2 (Supplementary Fig. 1), which were identified against the S177E/S181E (EE) mutant, in co-crystallization. A hIKK $\beta$  EE construct (1–678) lacking only the C-terminal NBD did crystallize; however, these crystals only diffracted to a resolution of  $\sim 7.5$  Å. Searching IKK $\beta$  orthologues that may give better crystals led to success in crystallizing the analogous region of xIKK $\beta$  EE (4–675; Fig. 1a).

The xIKK $\beta$  (S177E/S181E) protein containing residues 4–675 was concentrated by ultrafiltration (Amicon) to about  $15 \text{ mg ml}^{-1}$  in 20 mM Tris-HCl (pH 8.0), 150 mM NaCl and 10 mM dithiothreitol (DTT). It was mixed with an inhibitor compound in a 1:2 molar ratio before crystallization. Cmpd1 is 4-((4-(4-chlorophenyl)pyrimidin-2-yl)amino)phenyl(4-(2-hydroxyethyl)piperazin-1-yl)methanone and Cmpd2 is 1-(4-(4-((4-(pyridin-4-ylsulfonyl)phenyl)pyrimidin-2-yl)amino)benzoyl)piperazin-1-yl)ethanone. The P1 crystals were grown using hanging-drop vapour diffusion at 4 °C by mixing equal volumes of the purified protein and the crystallization condition of 100 mM *N*-(2-acetamido)iminodiacetic acid at pH 6.5, 10% (w/v) polyethylene glycol 6000, 50 mM Li<sub>2</sub>SO<sub>4</sub>, 300 mM NaCl and 10 mM DTT. The I4<sub>1</sub>22 crystals were grown at 4 °C with well solution containing 1.8 M K/Na phosphate at pH 5.6 and 10 mM DTT. For data collection, all crystals were flash frozen in the respective crystallization conditions supplemented with 25% (v/v) ethylene glycol. Diffraction data were collected at the 24ID-C beam line of the Advanced Photon Source. Multiwavelength anomalous diffraction (MAD) data on heavy-atom derivative crystals or selenomethionyl crystals were collected near the respective absorption edges. All diffraction data were processed using the HKL2000 suite<sup>43</sup> and their statistics are shown in Supplementary Table 1 and Supplementary Table 2.

**Structure determination, refinement and analysis.** The initial xIKK $\beta$  crystals grew in the P1 space group in the presence of the inhibitor Cmpd1 or Cmpd2 and diffracted to a resolution of 3.6 Å. Selenomethionyl crystals were obtained, but we failed to locate the large number of expected selenium sites. Among the extensive heavy-atom searches, an ytterbium-derivative was obtained, with eight well-defined sites, which probably correspond to eight IKK $\beta$  molecules in the asymmetric unit. However, the electron density map calculated from a three-wavelength ytterbium-anomalous diffraction data set was insufficient for tracing, and phase combination with the selenomethionyl data set could not be performed, owing to non-isomorphism.

The structure determination was eventually successful in the alternative crystal form, I4<sub>1</sub>22, which contains one molecule of IKK $\beta$  in complex with Cmpd1 and diffracted to a resolution of 4.0 Å, using MAD of the selenomethionyl crystals (Supplementary Tables 1 and 2 and Supplementary Fig. 2). Twelve selenium sites were determined using the program SHELXD<sup>44</sup> and refined with the program MLPHARE in the CCP4 suite<sup>45</sup>. MAD phases were calculated at a resolution of 4.0 Å with data from I4<sub>1</sub>22 crystals using the program SHARP<sup>46</sup>. A few cycles of model building and refinement were carried out with the program O<sup>47</sup> and REFMAC with TLS parameterization<sup>48</sup>. The I4<sub>1</sub>22 crystals contain one monomer in the asymmetric unit and 80% solvent when calculated with the entire IKK $\beta$  construct and 84% solvent when considering only the ordered part of the structure. The inhibitor Cmpd1 has density in the MAD experimental map and the  $F_o - F_c$

omit map. The Dundee PRODRG2 Server was used to generate topology and restraint files of the compound for refinement. The refined model contains residues 16–236, 243–286, 290–376, 384–394, 401–475 and 528–637 and Cmpd1.

The structure of the P1 form was determined by molecular replacement using the refined model of the I4<sub>1</sub>22 crystal form as the search model, in which eight molecules were located. Selenium sites of the single-wavelength anomalous diffraction data set of a selenomethionyl crystal in the P1 space group was calculated by difference Fourier analysis and used to verify residue registration in the P1 structure. Refinement in the P1 structure was conducted at a resolution of 3.6 Å and incorporated tight non-crystallographic symmetry restraints (root mean squared deviation of 0.02 Å in atom positions). After several rounds of refinement at a resolution of 3.6 Å, new electron densities appeared in the P1 crystal form to complete the model building. Although Cmpd2 was in the crystallization condition, it did not have clear density and was not included in the refinement. The refined model of the P1 crystal form contains four IKK $\beta$  dimers in the asymmetric unit. Three of the dimers encompass residues 16–236, 243–286, 290–376, 384–394, 401–551 and 559–666. One dimer contains the same residues as the structure in I4<sub>1</sub>22. The structures were analysed using the CCP4 suite<sup>45</sup> and the Dali server<sup>49</sup>, and the figures were made using PYMOL<sup>50</sup>.

**GST pull-down.** The tagged proteins were first purified with glutathione or Ni-NTA beads and their expression levels were assessed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Beads containing estimated equivalent quantities of the tagged proteins were mixed with the cell lysates or the purified versions of the interaction partners. The mixtures were incubated at room temperature (20 °C) for 1 h with rotation. After centrifugation, the supernatants were removed. The beads were then washed twice, eluted and subjected to SDS–PAGE analysis. All pull-down experiments were repeated two to four times with consistency.

**Kinase assays using anti-phospho-IkB $\alpha$  antibody.** The hIKK $\beta$  proteins ( $0.1 \mu\text{g} \mu\text{l}^{-1}$ ) were incubated with recombinant IkB $\alpha$  ( $1 \mu\text{g} \mu\text{l}^{-1}$ ) in 50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 2 mM DTT for 30 min at 30 °C. SDS–PAGE sample buffer was used to terminate the reactions. The products were separated on 15% SDS–PAGE and transferred to PVDF membranes. Anti-phospho-IkB $\alpha$  antibody (Cell Signaling Technology) was used to detect phospho-IkB $\alpha$ .

**Determination of  $K_m$  of hIKK $\beta$  for IkB $\alpha$  (1–54) and full-length IkB $\alpha$ .** To derive the  $K_m$  of hIKK $\beta$  for full-length IkB $\alpha$ , kinase assays were performed at substrate concentrations of 2.8, 5.3, 10.6, 21.3, 42.5 and 85  $\mu\text{M}$ . Similarly, to derive the  $K_m$  of hIKK $\beta$  for IkB $\alpha$  (1–54), kinase assays were performed at substrate concentrations of 5.3, 10.6, 21.3, 42.5, 85 and 170  $\mu\text{M}$ . A time course of the kinase reactions was first performed to select a hIKK $\beta$  amount and a time point within which the reactions are linear with time. The final selected reactions contain 10 ng baculovirus-expressed hIKK $\beta$ , 100 mM cold ATP and 1  $\mu\text{l}$  [ $\gamma$ -<sup>32</sup>P]ATP ( $3,000 \text{ Ci mmol}^{-1}$ , 1 mCi per 100  $\mu\text{l}$ ) in 25  $\mu\text{l}$  of reaction buffer containing 50 mM Tris-HCl at pH 8.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub> and 2 mM DTT. The phosphotransfer reaction was allowed to proceed for 10 min at 30 °C and quenched with SDS–PAGE sample buffer. The products were separated on 15% SDS–PAGE and subjected to autoradiography. The relative amounts of phosphorylated IkB $\alpha$  were quantified using IMAGEJ, plotted against total IkB $\alpha$  concentrations and fitted using nonlinear regression to the Michaelis–Menten equation to obtain  $K_m$  using SIGMAPLOT.

**Transfection, immunoprecipitation and kinase assay.** The constructs Flag-hIKK $\beta$  EE and its truncation mutants; HA-hIKK $\beta$  EE and its dimerization mutants L654D/W655D, W655D/L658D and L654D/W655D/L658D; and Flag-hIKK $\beta$  and its dimerization mutants L654D/W655D and W655D/L658D were generated in the vector pcDNA3 by conventional PCR. All IKK $\beta$  constructs were transfected in HEK293T cells with Lipofectamine 2000 (Invitrogen). After 24 h, cell extracts were immunoprecipitated with anti-Flag antibodies bound to agarose beads (M2, Sigma) or anti-HA bound to agarose beads (Sigma). IKK $\beta$  kinase assays were essentially done as previously described<sup>6,13</sup>. Briefly, immunoprecipitates were incubated with 2  $\mu\text{M}$  full-length IkB $\alpha$  (1–317) in 20 mM HEPES at pH 7.5, 10 mM MgCl<sub>2</sub>, 20 mM  $\beta$ -glycerophosphate, 10 mM PNPP, 50 mM Na<sub>3</sub>VO<sub>4</sub>, 1 mM DTT, 20 mM ATP, and 1–10 mCi [ $\gamma$ -<sup>32</sup>P]ATP at 30 °C for 30 min, and subjected to SDS–PAGE and autoradiography. Immunoblotting was performed using anti-Flag (Sigma), anti-HA (Sigma) or anti-IKK $\beta$  antibodies (Upstate, 05-535).

**Equilibrium analytical ultracentrifugation measurements.** Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter), using six-cell centre pieces with straight walls, a 12-mm path length and sapphire windows. Samples were kept and diluted in 50 mM Tris-HCl at pH 8.0 and 300 mM NaCl. Samples from wild-type protein were diluted to 6.9, 4.5 and 2.4  $\mu\text{M}$ , mutant L654D/W655D to 7.4, 4.8 and 2.6  $\mu\text{M}$  and mutant W655D/L658D to 4.9, 3.2 and 1.7  $\mu\text{M}$  for channels A, B and C, respectively. Dilution buffer was used as blank. All samples were run at 4 °C at 9,000 r.p.m. (5,900g; held for 20 h then scanned four times at 1-h intervals), 11,000 r.p.m. (8,800g; held for 10 h then scanned four times at 1-h intervals), 14,000 r.p.m. (14,300g; held for

10 h then scanned four times at 1-h intervals) and 17,000 r.p.m. (21,000g; held for 10 h then scanned four times at 1-h intervals). Detection was by ultraviolet absorption at 280 nm. Solvent density and the protein partial specific volume at each temperature were determined using the program SEDNTERP (Alliance Protein Laboratories). For calculation of  $K_D$  and the apparent molecular weight, all useful data were used in a global fit, using the program HETEROANALYSIS, obtained from University of Connecticut (<http://www.biotech.uconn.edu/auf>).

43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
45. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
46. Bricogne, G. *et al.* Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
47. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building models in electron density maps and the location of errors in those models. *Acta Crystallogr. A* **47**, 110–119 (1991).
48. Winn, M. D., Murshudov, G. N. & Papiz, M. Z. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.* **374**, 300–321 (2003).
49. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).
50. DeLano, W. L. PyMOL Molecular Viewer (<http://www.pymol.org>) (2002).



# Structure and mechanism of the uracil transporter UraA

Feiran Lu<sup>1\*</sup>, Shuo Li<sup>1\*</sup>, Yang Jiang<sup>1</sup>, Jing Jiang<sup>1</sup>, He Fan<sup>1</sup>, Guifeng Lu<sup>1</sup>, Dong Deng<sup>1</sup>, Shangyu Dang<sup>1</sup>, Xu Zhang<sup>1</sup>, Jiawei Wang<sup>1</sup> & Nieng Yan<sup>1</sup>

**The nucleobase/ascorbate transporter (NAT) proteins, also known as nucleobase/cation symporter 2 (NCS2) proteins, are responsible for the uptake of nucleobases in all kingdoms of life and for the transport of vitamin C in mammals<sup>1,2</sup>. Despite functional characterization of the NAT family members in bacteria, fungi and mammals, detailed structural information remains unavailable. Here we report the crystal structure of a representative NAT protein, the *Escherichia coli* uracil/H<sup>+</sup> symporter UraA, in complex with uracil at a resolution of 2.8 Å. UraA has a novel structural fold, with 14 transmembrane segments (TMs) divided into two inverted repeats. A pair of antiparallel  $\beta$ -strands is located between TM3 and TM10 and has an important role in structural organization and substrate recognition. The structure is spatially arranged into a core domain and a gate domain. Uracil, located at the interface between the two domains, is coordinated mainly by residues from the core domain. Structural analysis suggests that alternating access of the substrate may be achieved through conformational changes of the gate domain.**

Representative NAT/NCS2 family proteins include the vitamin C transporters SVCT1 and SVCT2 in mammals<sup>3–5</sup>; the uracil transporter SNBT1 in rat<sup>6</sup>; the uric acid/xanthine transporter UapA in *Aspergillus nidulans*<sup>7</sup>; the UapA homologue, YgfO (also known as XanQ), in *E. coli*<sup>8</sup>; and the uracil and 5-fluorouracil transporter UraA<sup>9</sup> (Supplementary Fig. 1). NAT proteins contain a signature motif, (A, G, S)(Q, E, P)N-X-G-X<sub>4</sub>-T(R, K, G), where X denotes a nonspecific amino acid, that may be involved in substrate recognition and transport<sup>10–12</sup>. Whereas the mammalian NAT proteins are sodium symporters<sup>6</sup>, their bacterial homologues are proton symporters<sup>2</sup>. Uptake of nucleobases can also be mediated by members of the nucleobase/cation symporter 1 (NCS1) family<sup>1,13</sup>, which share little sequence similarity with NAT/NCS2 proteins. The NCS1 transporter Mhp1 is structurally similar to LeuT<sup>14–16</sup>. It remains to be seen whether the NAT/NCS2 transporters also conform to the LeuT fold.

Uracil binds to recombinant UraA protein with a dissociation constant of approximately  $0.41 \pm 0.07 \mu\text{M}$  (s.d. of three independent experiments) as measured by scintillation proximity assay<sup>17</sup> (Supplementary Fig. 2a). Uracil proved to be essential for generation of usable UraA crystals grown in the presence of *n*-nonyl- $\beta$ -D-glucopyranoside ( $\beta$ -NG). We determined the structure using mercury-based single isomorphous replacement (Supplementary Table 1). The electron density was of excellent quality (Supplementary Fig. 3a) except for that of residues 179–195, which were modelled as poly-Ala. During structure refinement, a prominent, disc-shaped electron density reminiscent of uracil appeared in the centre of UraA after most amino acids were modelled (Supplementary Fig. 3b). Because crystal growth strictly depended on the inclusion of 1 mM uracil in the protein purification and crystallization solution, we modelled uracil into the density and refined the final atomic model at a resolution of 2.8 Å (Supplementary Table 1 and Supplementary Fig. 3c).

UraA contains 14 transmembrane segments, with both the amino and the carboxy termini located on the cytoplasmic side (Fig. 1a). The

two C-terminal  $\alpha$ -helices,  $\alpha$ 13 and  $\alpha$ 14, are only halfway into the membrane. There is no detectable structural similarity between UraA and Mhp1<sup>14</sup>. Searching the Protein Data Bank using the Dali server<sup>18</sup> suggested that UraA has a novel fold (Supplementary Fig. 4a). Notably, two short, antiparallel  $\beta$ -strands on TM3 and TM10 are located at the centre of the structure. Each  $\beta$ -strand is preceded by an extended, unwound fragment and followed by a short  $\alpha$ -helix,  $\alpha$ 3 or  $\alpha$ 10 (Fig. 1b). Discontinuous helices have been observed in a number of transport proteins<sup>19</sup>, but it is unusual for an unwound region to constitute half of the transmembrane segment. These structural features distinguish UraA from other known structures of integral membrane proteins. The 11-residue NAT signature motif, which was predicted to be a loop in the cytoplasm<sup>10,12</sup>, constitutes membrane-embedded  $\alpha$ 10 in UraA (Supplementary Fig. 1).

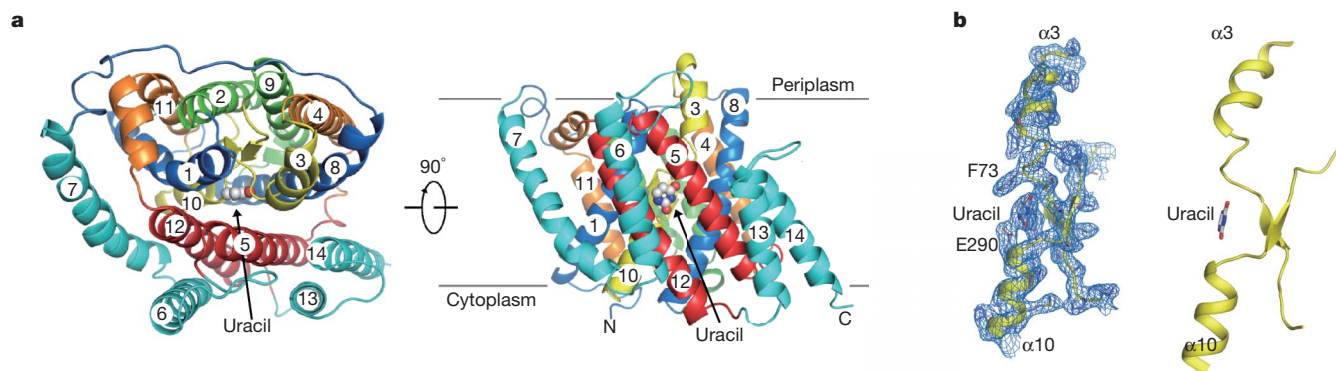
The 14 transmembrane segments of UraA are arranged into two structural repeats, TM1–TM7 and TM8–TM14, which are related to each other by a rotation of approximately 180° around an axis parallel to the membrane bilayer. These two repeats can be superimposed with a root mean squared deviation of 2.9 Å over 135 aligned C $\alpha$  atoms (Supplementary Fig. 4b). The presence of two inverted repeats is a shared feature for a number of transporters<sup>16</sup> and channels<sup>20,21</sup>.

The 14 transmembrane segments are spatially organized into a core domain and a gate domain. The core domain comprises TM1–TM4 and TM8–TM11, and the gate domain contains the other six segments (Fig. 2a, left). The interface between the two domains is populated mainly with hydrophobic residues from TM1, TM3 and TM8 in the core domain and TM5 and TM12 in the gate domain (Fig. 2a, right). Notably, the interdomain interactions are further stabilized by the head group of a bound  $\beta$ -NG molecule (Supplementary Fig. 5). We added  $\beta$ -NG in the last step of protein purification, whereas uracil was included throughout purification and crystallization. Uracil binds UraA in the absence of  $\beta$ -NG (Supplementary Fig. 2a), whereas addition of  $\beta$ -NG to substrate-free UraA led to severe protein precipitation that prevented biochemical analysis. Furthermore, D-glucose, the head group of  $\beta$ -NG, had no impact on uracil binding or transport when measured by scintillation proximity assay or cell-uptake assay. Thus, we concluded that uracil-bound UraA may provide a suitable conformation for  $\beta$ -NG to bind to and that  $\beta$ -NG binding is unlikely to have an impact on uracil recognition.

In contrast to the hydrophobic interface between the two domains, there are a large number of buried hydrogen bonds within the core domain, with the  $\beta$ -strands of TM3 and TM10 at the centre (Fig. 2b). Apart from conventional hydrogen bonds in antiparallel  $\beta$ -strands, the hydroxyl group of Tyr 288 in TM10 makes two hydrogen bonds with the carbonyl oxygen and amide nitrogen of Ser 71 in TM3. There are 12 additional hydrogen bonds between TM3/TM10 and other transmembrane segments (Fig. 2b). These extensive interactions may facilitate the conformation of the unwound regions in TM3 and TM10 as well as restraining movement of the transmembrane segments in the core domain.

<sup>1</sup>State Key Laboratory of Bio-membrane and Membrane Biotechnology, Center for Structural Biology, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China.

\*These authors contributed equally to this work.

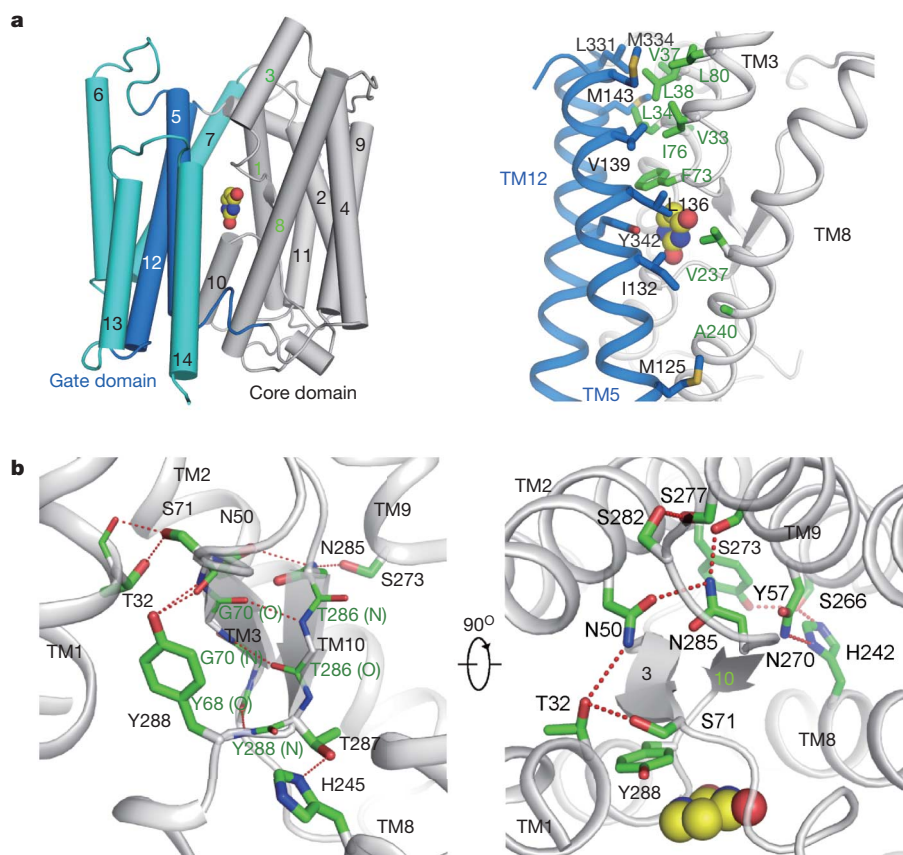


**Figure 1 | The structure of UraA reveals a novel fold.** **a**, Overall structure of UraA. Two perpendicular views, one from the periplasm and one from the side, are shown. The bound uracil is indicated. **b**, A pair of short, antiparallel  $\beta$ -strands is located in the middle of TM3 and TM10. The  $2F_o - F_c$  electron

density map, shown in blue mesh, is contoured at  $1\sigma$  on the left. All structural figures, including the calculation of surface electrostatic potential, were prepared with PYMOL<sup>30</sup>.

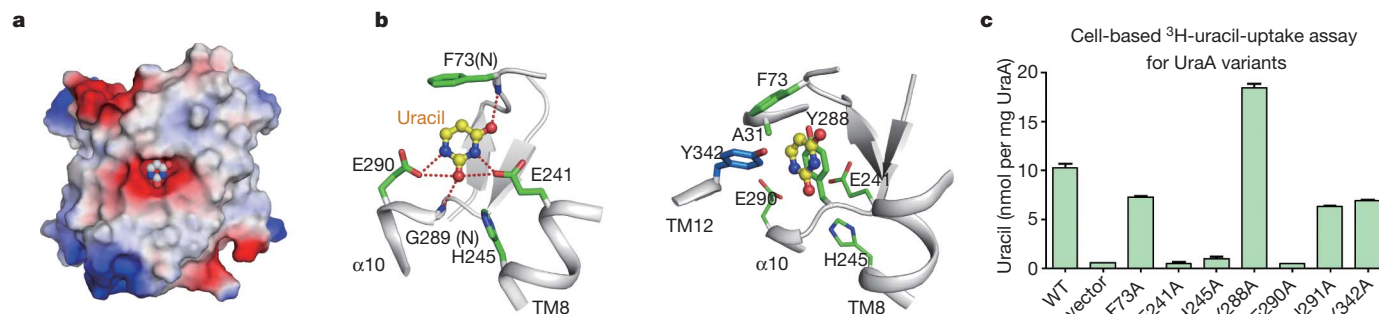
The antiparallel  $\beta$ -strands and the connecting loops of TM3 and TM10 provide a shelter for uracil between the core and gate domains. The pyrimidine ring of uracil, which is roughly parallel to the  $\beta$ -strands, is surrounded by negative electrostatic potentials (Fig. 3a). Recognition of uracil is almost exclusively mediated by the core domain, involving residues from TM1, TM3, TM8, TM10 and TM12 (Fig. 3b and Supplementary Fig. 4c). Two Glu residues, Glu 241 and Glu 290, anchor uracil by each making two hydrogen bonds with it (Fig. 3b). Replacement of either Glu residue by Ala completely abrogated uracil binding (Supplementary Fig. 2b). In addition, the two oxygen atoms of uracil

form hydrogen bonds with the amide nitrogen atoms of Phe 73 and Gly 289. Notably, Gly 289 and Glu 290 are within the NAT signature motif. A hydroxyl group in  $\beta$ -NG forms hydrogen bonds with both the imidazole nitrogen of His 245 and the carbonyl oxygen of uracil (Supplementary Fig. 5b), suggesting that a water molecule may occupy the position of the  $\beta$ -NG hydroxyl group and make similar interactions. In this case, His 245 may contribute to uracil binding with a water-mediated hydrogen bond (Supplementary Fig. 5c). Consistent with this notion, substitution of Ala for His 245 abolished uracil binding (Supplementary Fig. 2b).



**Figure 2 | Domain organization of UraA.** **a**, Left: UraA is spatially organized into core (grey) and the gate (cyan and blue) domains that associate with each other through hydrophobic interactions. Right: the residues that mediate the interdomain interactions are shown as green (core) and blue (gate) sticks in the right panel. Uracil is shown as yellow spheres. **b**, Antiparallel  $\beta$ -strands in TM3

and TM10 provide the organizing centre for the core domain. There are a large number of hydrogen bonds between transmembrane segments within the core domain. Two perpendicular views are shown. Hydrogen bonds are represented by red dashed lines.



**Figure 3 | Uracil coordination by UraA.** **a**, Uracil is located in a concave pocket surrounded by negative electrostatic potential. The core domain of UraA is shown in surface electrostatic potential. **b**, Uracil is coordinated by both polar (left) and van der Waals (right) contacts. Uracil is shown in yellow, in ball-and-stick form. **c**, Cell-based <sup>3</sup>H-uracil-uptake assay identified the key residues in uracil transport. The membrane expression levels of UraA variants

were monitored by western blot using an antibody against the His tag. The amount of each protein was estimated by comparing the intensity against a serial dilution of UraA with known concentration on the same western blot. The reactions lasted for 30 s. Details of the experiments are described in Supplementary Fig. 6 and Methods. WT, wild type. Error bars, s.d. of three independent experiments.

In addition to hydrogen bonds, uracil is coordinated by van der Waals interactions involving Ala 31, Phe 73, Tyr 288 and Tyr 342. Phe 73 blocks access to uracil from the periplasm, whereas the phenyl ring of Tyr 288 is roughly parallel to the pyrimidine ring of uracil. Tyr 342 is the only residue from the gate domain that contributes to uracil binding (Fig. 3b). The presence of aromatic residues surrounding the substrate is commonly observed in membrane transporters, such as the sodium-coupled sugar transporter vSGLT<sup>22</sup>, the glycine betaine transporter BetP<sup>23</sup> and the arginine/agmatine antiporter AdiC<sup>24,25</sup>. The bulky residues effectively insulate the substrate from the outside environment.

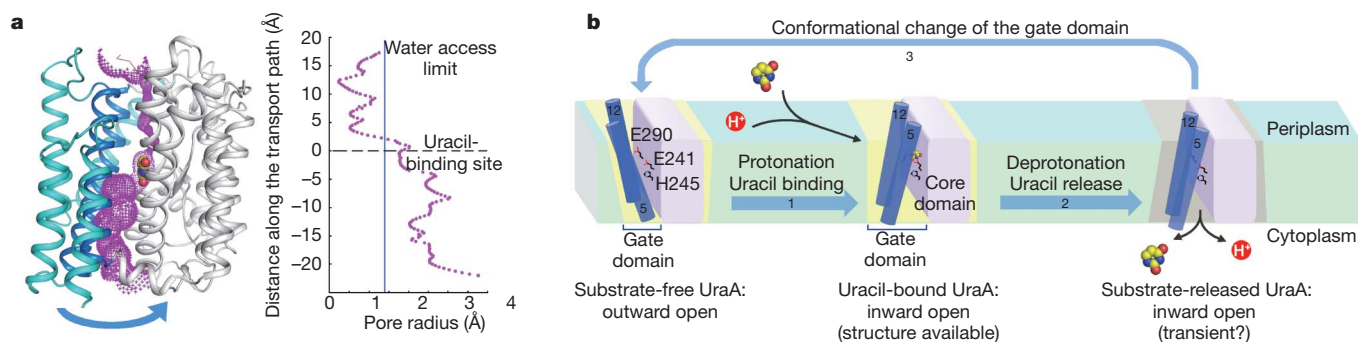
To corroborate structural observations, we generated a number of UraA variants, each containing a mis-sense mutation, and examined their ability to transport uracil into *E. coli*. Whereas the wild-type UraA can mediate the uptake of <sup>3</sup>H-uracil with a Michaelis constant,  $K_m$ , of approximately 0.5  $\mu$ M (Supplementary Fig. 6), replacement of Glu 241, His 245 or Glu 290 by Ala invariably abrogated uracil uptake (Fig. 3c). This finding is consistent with loss of uracil binding by these three mutants (Supplementary Fig. 2). By contrast, the UraA variants Phe 73 Ala and Tyr 342 Ala retained the bulk of transport activities (Fig. 3c).

Mutation of the invariant Asn residue in the NAT motif was shown to abrogate the transport function for UapA and YgfO<sup>10–12</sup>. In UraA, however, only Gly 289 and Glu 290 of the NAT motif are directly involved in substrate binding. Asn 291 is located away from the substrate-binding site (Supplementary Fig. 7). The UraA variant Asn 291 Ala retained the majority of the wild-type uracil-uptake activity (Fig. 3c). Because UapA

and YgfO are xanthine/uric acid permease, the invariant Asn residue in the NAT motif may have a more important role in these purine transporters.

UraA is a proton-coupled symporter. Translocation of proton relies on the residues that can be protonated and deprotonated. Remarkably, Glu 241, His 245 and Glu 290, all of which can undergo cycles of protonation and deprotonation, are clustered at the interface between the core and gate domains (Supplementary Fig. 8) and are essential for uracil binding (Fig. 3b and Supplementary Fig. 2b). This arrangement suggests that these residues may have a key role in proton translocation. Curiously, however, only Glu 241 is invariant among all NAT members. His 245 is replaced by Asp, Thr or Val in other NAT proteins, whereas Glu 290 is substituted by Gln or Pro (Supplementary Fig. 1). A closer examination revealed a pattern of function-based conservation. Glu 290 is conserved in the three known pyrimidine permeases RutG, PyrP and SNBT1. His 245 is conserved in RutG and PyrP, both of which are pyrimidine/proton symporters. By contrast, His 245 is replaced by a conserved Asp residue in the sodium symporters SNBT1, SVCT1 and SVCT2 (Supplementary Fig. 1). This analysis suggests conserved mechanisms of transport, with His 245 having a key role in proton symport and Asp having a similar role in sodium symport.

An analysis of solvent accessibility using the program HOLE<sup>26</sup> unambiguously shows that uracil is occluded from the periplasm but readily accessible from the cytoplasmic side (Fig. 4a). Thus, our current structure of uracil-bound UraA adopts an inward-open conformation. To load uracil, UraA must also be able to adopt an outward-open conformation. This is probably accomplished by a rigid-body rotation



**Figure 4 | Working model for the transport mechanism of UraA.** **a**, Uracil-bound UraA is in an inward-open conformation. The van der Waals surface of UraA (purple) was calculated with the program HOLE<sup>26</sup>, which revealed that the bound uracil is insulated from the periplasm but exposed to the cytoplasm. The pore radii along the potential transport path are tabulated on the right.

**b**, Working model to illustrate the putative transport mechanism of UraA. Glu 241, His 245 and Glu 290 are shown to emphasize their essential role in substrate recognition and proton translocation. For simplicity, only TM5 and TM12 are shown to represent the gate domain.



of the gate domain relative to the core domain around the bound uracil (Fig. 4a).

On the basis of structural and biochemical analyses, we propose a working model to explain proton-coupling and uracil symport by UraA (Fig. 4b). The default conformation of substrate-free UraA may be outward-open, with the two Glu residues deprotonated. The negative charges on the Glu residues make closure of the gate domain onto the core domain energetically unfavourable. On binding of uracil and  $H^+$ , at least one Glu residue is protonated and the gate domain undergoes a conformational change, taking an inward-open conformation as seen in the structure. A proton translocation is likely to occur from Glu to His 245 and onward into the cytoplasm. Deprotonation may cause local conformational changes around Glu 241, His 245 and Glu 290, leading to the dissociation of uracil. An inward-open, deprotonated conformation is likely to be transient and quickly reverts to the outward-open state. In the sodium symporters of the NAT family, the sodium ion, rather than the proton, is required to neutralize the conserved Glu and Asp residues and for substrate binding.

This model predicts that the core domain provides the molecular basis of substrate selectivity and proton/sodium translocation, whereas conformational changes of the gate domain allow transport of substrate. We recognize the speculative nature of this model, as many important questions remain unanswered and require experimental investigation. For example, we do not know the molar ratio between the proton/sodium ion and the substrate molecule during each transport cycle, nor do we know how the conserved Glu and His residues trigger conformational changes during protonation and deprotonation. Nonetheless, the structural and biochemical characterizations of UraA reported here provide an important framework for mechanistic understanding of the NAT family transporters.

## METHODS SUMMARY

We generated all clones using a standard PCR-based cloning strategy. Wild-type and mutant UraA were expressed and purified to homogeneity. We grew crystals of wild-type UraA by the hanging-drop vapour diffusion method in the presence of 0.4%  $\beta$ -NG (Anatrace). All data were collected at SPring-8 beamline BL41XU and processed with the HKL2000 package<sup>27</sup>. The structure was determined by Hg-SIRAS using SOLVE<sup>28</sup> and refined with PHENIX<sup>29</sup>. We measured the binding affinity between uracil and purified recombinant UraA by scintillation proximity assay<sup>17</sup>. The cell-based uracil-uptake assay was modified on the basis of a published protocol<sup>19</sup>.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 23 September 2010; accepted 27 January 2011.

Published online 20 March 2011.

1. Saier, M. H. Jr *et al.* Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochim. Biophys. Acta* **1422**, 1–56 (1999).
2. Gournas, C., Papageorgiou, I. & Dhalluin, G. The nucleobase-ascorbate transporter (NAT) family: genomics, evolution, structure-function relationships and physiological role. *Mol. Biosyst.* **4**, 404–416 (2008).
3. Tsukaguchi, H. *et al.* A family of mammalian  $Na^+$ -dependent L-ascorbic acid transporters. *Nature* **399**, 70–75 (1999).
4. Wang, Y. *et al.* Human vitamin C (L-ascorbic acid) transporter SVCT1. *Biochem. Biophys. Res. Commun.* **267**, 488–494 (2000).
5. Savini, I., Rossi, A., Pierro, C., Avigliano, L. & Catani, M. V. SVCT1 and SVCT2: key proteins for vitamin C uptake. *Amino Acids* **34**, 347–355 (2008).
6. Yamamoto, S. *et al.* Identification and functional characterization of the first nucleobase transporter in mammals: implication in the species difference in the intestinal absorption mechanism of nucleobases and their analogs between higher primates and other mammals. *J. Biol. Chem.* **285**, 6522–6531 (2010).
7. Dhalluin, G. & Scazzocchio, C. A gene coding for the uric acid-xanthine permease of *Aspergillus nidulans*: inactivation cloning, characterization, and sequence of a cis-acting mutation. *Genetics* **122**, 341–350 (1989).

8. Karatz, P. & Frillingos, S. Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*. *Mol. Membr. Biol.* **22**, 251–261 (2005).
9. Andersen, P. S., Frees, D., Fast, R. & Mygind, B. Uracil uptake in *Escherichia coli* K-12: isolation of *uraA* mutants and cloning of the gene. *J. Bacteriol.* **177**, 2008–2013 (1995).
10. Koukaki, M. *et al.* The nucleobase-ascorbate transporter (NAT) signature motif in UraA defines the function of the purine translocation pathway. *J. Mol. Biol.* **350**, 499–513 (2005).
11. Karatz, P., Panos, P., Georgopoulou, E. & Frillingos, S. Cysteine-scanning analysis of the nucleobase-ascorbate transporter signature motif in YgfO permease of *Escherichia coli*: Gln-324 and Asn-325 are essential, and Ile-329–Val-339 form an  $\alpha$ -helix. *J. Biol. Chem.* **281**, 39881–39890 (2006).
12. Georgopoulou, E., Mermelekas, G., Karena, E. & Frillingos, S. Purine substrate recognition by the nucleobase-ascorbate transporter signature motif in the YgfO xanthine permease: Asn-325 binds and Ala-323 senses substrate. *J. Biol. Chem.* **285**, 19422–19433 (2010).
13. de Koning, H. & Dhalluin, G. Nucleobase transporters. *Mol. Membr. Biol.* **17**, 75–94 (2000).
14. Weyand, S. *et al.* Structure and molecular mechanism of a nucleobase-cation-symport-1 family transporter. *Science* **322**, 709–713 (2008).
15. Shimamura, T. *et al.* Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science* **328**, 470–473 (2010).
16. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of  $Na^+/Cl^-$ -dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
17. Quick, M. *et al.* Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc. Natl. Acad. Sci. USA* **106**, 5563–5568 (2009).
18. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
19. Screpanti, E. & Hunte, C. Discontinuous membrane helices in transport proteins and their correlation with function. *J. Struct. Biol.* **159**, 261–267 (2007).
20. Fu, D. *et al.* Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* **290**, 481–486 (2000).
21. Wang, Y. *et al.* Structure of the formate transporter FocA reveals a pentameric aquaporin-like channel. *Nature* **462**, 467–472 (2009).
22. Faham, S. *et al.* The crystal structure of a sodium galactose transporter reveals mechanistic insights into  $Na^+$ /sugar symport. *Science* **321**, 810–814 (2008).
23. Ressler, S., Terwisscha van Scheltinga, A. C., Vonrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the  $Na^+$ /betaine symporter BetP. *Nature* **458**, 47–52 (2009).
24. Fang, Y. *et al.* Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature* **460**, 1040–1043 (2009).
25. Gao, X. *et al.* Mechanism of substrate recognition and transport by an amino acid antiporter. *Nature* **463**, 828–832 (2010).
26. Smart, O. S., Goodfellow, J. M. & Wallace, B. A. The pore dimensions of gramicidin A. *Biophys. J.* **65**, 2455–2460 (1993).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
29. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
30. DeLano, W. L. PyMOL Molecular Viewer (<http://www.pymol.org>) (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank N. Shimizu at the Spring-8 beamline BL41XU, and J. He and S. Huang at the Shanghai Synchrotron Radiation Facility for on-site assistance. We thank Y. Shi and Y. Yan for critical discussions. This work was supported by funds from the Ministry of Science and Technology (grant numbers 2009CB918802 and 2011CB910501), Tsinghua University 985 Phase II funds and Project 91017011 supported by the National Natural Science Foundation of China. N.Y. acknowledges support from the Yuyuan Foundation and the Li Foundation.

**Author Contributions** F.L., S.L., Y.J., J.J., H.F., G.L. and N.Y. designed all experiments. F.L., S.L., Y.J., J.J., H.F., G.L., D.D., S.D. and X.Z. performed the experiments. All authors analysed the data. F.L., J.W. and N.Y. contributed to manuscript preparation. N.Y. wrote the manuscript.

**Author Information** The atomic coordinates of UraA have been deposited in the Protein Data Bank under accession code 3QE7. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to N.Y. ([nyan@tsinghua.edu.cn](mailto:nyan@tsinghua.edu.cn)).

## METHODS

**Protein preparation.** The complementary DNA of full-length UraA from *E. coli* strain O157:H7 was subcloned into pET21b vector (Novagen). The UraA mutants were generated using two-step PCR and subcloned, overexpressed and purified in the same way as wild-type protein. Overexpression of UraA was induced in *E. coli* BL21(DE3) by 0.2 mM isopropyl- $\beta$ -D-thiogalactoside (IPTG) when the cell density reached  $D_{600\text{ nm}}$  1.5. To obtain the structure of uracil-bound UraA, uracil (Sigma) was added at 1 mM when cells were induced. Uracil (1 mM) was included in all the buffers during protein purification. After growth for 16 h at 22 °C, the cells were collected, resuspended in buffer containing 25 mM Tris-HCl, pH 8.0, and 150 mM NaCl, and disrupted using a French press with two passes at 10,000–15,000 p.s.i. Cell debris was removed by centrifugation at 27,000g for 10 min. The supernatant containing the membrane was collected and underwent ultracentrifugation at 150,000g for 1 h. The membrane fraction in the pellet was harvested and incubated with 1.5% (w/v) *n*-dodecyl- $\beta$ -D-maltopyranoside (DDM, Anatrace) for 1 h at 4 °C. After another ultracentrifugation step at 150,000g for 30 min, the supernatant was collected and loaded on Ni<sup>2+</sup>-nitrilotriacetate affinity resin (Ni-NTA, Qiagen). Subsequently, the resin was washed three times, each time with 10 ml buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 30 mM imidazole and 0.02% DDM. The protein was eluted from the affinity resin with 10 ml wash buffer plus 250 mM imidazole. The proteins were concentrated to about 10 mg ml<sup>-1</sup> before undergoing gel-filtration chromatography (Superdex-200 10/30, GE Healthcare) in buffer containing 25 mM Tris-HCl, pH 8.0, 150 mM NaCl and the indicated detergents. The peak fractions were collected for biochemical analyses or crystallization trials.

**Crystallization.** Crystals were grown at 18 °C by the hanging-drop vapour diffusion method. Crystals in the space group *P*6<sub>4</sub>22 were obtained for protein purified in the presence of 0.4%  $\beta$ -NG. Uracil (1 mM) was included during protein expression, purification and crystallization. Crystals appeared overnight in the buffer containing 25% PEG400, 100 mM MES-NaOH, pH 6.5, and 300 mM Li<sub>2</sub>SO<sub>4</sub>, and typically grew to form 50  $\mu$ m  $\times$  50  $\mu$ m  $\times$  100  $\mu$ m hexagonal rods in about one month. The crystals diffracted X-rays beyond 2.9 Å at SPring-8 beamline BL41XU. Mercury derivatives were obtained by soaking the crystals for 24 h in the mother liquor containing 1 mg ml<sup>-1</sup> (C<sub>2</sub>H<sub>5</sub>HgO)HPO<sub>2</sub>. Both native and heavy-atom-derived crystals were directly flash frozen in a cold nitrogen stream at 100 K.

**Data collection and structure determination.** All the data were collected at SPring-8 beamline BL41XU and processed with the HKL2000 package<sup>27</sup>. Further processing was carried out with programs from the CCP4 suite<sup>31</sup>. Data collection statistics are summarized in Supplementary Table 1. The initial phases were obtained from the native *P*6<sub>4</sub>22 crystal and its Hg derivative by single isomorphous replacement with anomalous scattering (SIRAS) using the program SOLVE<sup>28</sup>. Figures of merit before and after density modification are 0.244 and 0.749, respectively. The real-space constraints were applied to the electron density map in DM<sup>32</sup>. The model was built manually in COOT<sup>33</sup> and the structure was refined with PHENIX<sup>29</sup>. All structure figures, including the calculation of surface electrostatic potential, were prepared with PYMOL<sup>30</sup>.

**Cell-based uracil-uptake assay.** The *uraA*-deficient *E. coli* strain Keio Collection JW2482 (*F*-,  $\Delta$ (*araD-araB*)567,  $\Delta$ (*lacZ4787*::*rrnB*-3),  $\phi$ -*lambda*-,  $\Delta$ (*uraA745::kan*, *rph*-1,  $\Delta$ (*rhaD-rhaB*)568, *hsdR514*) used in this assay was purchased from National BioResource Project (Japan). Wild-type and UraA mutants were subcloned into pQLINK vector<sup>34</sup> with His<sub>6</sub> tag at the C terminus. The membrane expression levels of UraA variants were monitored by western blot using an antibody against the His tag. The amount of each UraA protein in the membrane fraction was estimated by comparing the intensity of the proteins in the membrane

fraction against a serial dilution of purified UraA with known concentrations on the same western blot, a protocol reported previously<sup>35</sup>.

The cell-based uptake assay was performed following the published protocol with minor modifications<sup>9</sup>. *uraA*-deficient *E. coli* cells transformed with plasmids were grown in LB medium at 37 °C and induced with 50  $\mu$ M IPTG when the cell density reached  $D_{600\text{ nm}}$  1.5. Cells were collected after induction for 30 min, rinsed and resuspended to an adjusted  $D_{600\text{ nm}}$  2.0 in AB medium (a modified minimal medium; see [http://openwetware.org/wiki/AB\\_medium](http://openwetware.org/wiki/AB_medium) for the detailed recipe). After incubation in AB medium for 1 h, the cells were taken for the uracil-uptake assay.

For time-course experiments, [5,6-<sup>3</sup>H]-uracil (4 Ci mmol<sup>-1</sup>, American Radiolabelled Chemicals) was added at indicated concentrations. All the reactions were performed at 25 °C. At the indicated reaction time, an aliquot of cells was taken for filtration through 0.45- $\mu$ m cellulose acetate filter (Sartorius). The filter membranes were immediately washed with 2 ml ice-cold AB medium, dried and taken for liquid scintillation counting. Control experiments were performed with cells transformed with pQLINK empty vector. For each data point, a parallel control experiment was performed and the control flux was subtracted before data fitting.

The time-course experiments showed that the accumulation of uracil was roughly linear within the first 30–60 s. Therefore, to determine the  $K_m$  and  $V_{\text{max}}$  of uracil uptake by wild-type UraA, the initial velocities were measured at 30 s. All experiments, including control experiments, were repeated at least three times, and the data were fitted to the Michaelis-Menten equation,  $V = V_{\text{max}}[\text{UraA}]/(K_m + [\text{UraA}])$ , in GRAPHPAD PRISM 5.0 Demo.

To compare the transport activity of the UraA variants, [5,6-<sup>3</sup>H]-uracil (40 Ci mmol<sup>-1</sup>) was applied at 0.25  $\mu$ M and each reaction was allowed for 30 s. All the UraA variants were expressed and quantified following the same protocol as for wild-type UraA. The solution behaviour of the purified UraA variants was examined by size exclusion chromatography, which showed similar profiles to that of wild-type protein.

**SPA-based binding assay.** Scintillation proximity assay (SPA) beads were diluted to 2 mg ml<sup>-1</sup> in 150 mM MES-NaOH, pH 6.4, 50 mM NaCl, 20% glycerol, 2 mM TCEP (Sigma) and 0.05% DDM. About 400 ng of purified, C-terminal His<sub>6</sub>-tagged UraA protein was incubated with [5,6-<sup>3</sup>H]-uracil at 4 °C for 2 h. The solution was then added to 100  $\mu$ l SPA beads and incubated by vigorous shaking at 4 °C in the dark for 2 h. The mixture was then loaded into individual wells of 96-well plates. Scintillation was read in the SPA mode on a Wallac 1450 MicroBeta plate PMT counter. For the competition assay, the final concentration of [5,6-<sup>3</sup>H]-uracil was kept constant at 0.17  $\mu$ M while the concentration of non-labelled uracil was increased from 0 to 10  $\mu$ M. To define the nonspecific binding activity, 400 mM imidazole was added to the wells to compete with His<sub>6</sub> for bead binding. All experiments were performed at least three times and data are presented as mean  $\pm$  s.d. Nonspecific binding was subtracted from each data point. Data fitting was performed using GRAPHPAD PRISM 5.0 Demo.

31. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
32. Cowtan, K. DM: an automated procedure for phase improvement by density modification. *Joint CCP4 ESF-EACBM Newslett. Protein Crystallogr.* **31**, 34–38 (1994).
33. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
34. Scheich, C., Kummel, D., Soumailakakis, D., Heinemann, U. & Bussow, K. Vectors for co-expression of an unrestricted number of proteins. *Nucleic Acids Res.* **35**, e43 (2007).
35. Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738 (2010).

# Redox freezing and melting in the Earth's deep mantle resulting from carbon–iron redox coupling

Arno Rohrbach<sup>1</sup> & Max W. Schmidt<sup>1</sup>

Very low seismic velocity anomalies in the Earth's mantle<sup>1,2</sup> may reflect small amounts of melt present in the peridotite matrix, and the onset of melting in the Earth's upper mantle is likely to be triggered by the presence of small amounts of carbonate<sup>3</sup>. Such carbonates stem from subducted oceanic lithosphere in part buried to depths below the 660-kilometre discontinuity and remixed into the mantle. Here we demonstrate that carbonate-induced melting may occur in deeply subducted lithosphere at near-adiabatic temperatures in the Earth's transition zone and lower mantle. We show experimentally that these carbonatite melts are unstable when infiltrating ambient mantle and are reduced to immobile diamond when recycled at depths greater than ~250 kilometres, where mantle redox conditions are determined by the presence of an (Fe,Ni) metal phase<sup>4–6</sup>. This 'redox freezing' process leads to diamond-enriched mantle domains in which the Fe<sup>0</sup>, resulting from Fe<sup>2+</sup> disproportionation in perovskites and garnet, is consumed but the Fe<sup>3+</sup> preserved. When such carbon-enriched mantle heterogeneities become part of the upwelling mantle, diamond will inevitably react with the Fe<sup>3+</sup> leading to true carbonatite redox melting at ~660 and ~250 kilometres depth to form deep-seated melts in the Earth's mantle.

The Earth's mantle is a heterogeneous, marble-cake-like composite of pristine, depleted, and mostly pyroxenitic or eclogitic recycled material<sup>7,8</sup>, further complicated by heterogeneities left behind in ancient melt passageways (for example, dunites, cumulative pyroxenites, or carbon-enriched domains formed through redox freezing). These heterogeneities are different from average mantle in terms of bulk composition<sup>7</sup>, or formed through differences in H<sub>2</sub>O or CO<sub>2</sub> volatile influx. Melting processes in the deep mantle are triggered by such heterogeneities, as solidus temperatures for fertile or depleted peridotite are significantly higher than adiabatic temperatures at depths >120 km. Using high pressure experiments, we investigate the formation of C-enriched mantle domains in the vicinity of deeply subducted lithosphere (through redox reactions between recycled carbonate and ambient metal-bearing mantle) and the remelting of such domains when entrained into the convective mantle.

Carbonated peridotite systems are central to understanding processes that involve the recycling of subducted, carbonated lithosphere back into the mantle. Experiments have shown that at pressures above ~2.5 GPa, carbonated mantle melts at lower temperatures than carbonate-free mantle<sup>3,9–12</sup>. However, these experimental studies have maintained carbon in its oxidized form. To understand what happens within the generally reduced deep Earth, it is necessary to include redox equilibria (controlled by oxygen fugacity,  $f_{O_2}$ ) between intrinsically oxidized carbonate minerals or melts and the reduced and metal-bearing deep mantle.

The redox state of the mantle determines whether carbon is present in its oxidized and potentially mobile form as carbonate or carbonatite melt (which lower the mantle solidus by several hundred degrees<sup>3,9–12</sup>), or whether it is present in its reduced and immobile form as graphite or diamond (which do not affect melting temperatures). In the uppermost mantle to ~250 km depth,  $f_{O_2}$  is determined by Fe<sup>2+</sup>/Fe<sup>3+</sup> equilibria in silicate minerals. Studies of natural peridotite xenoliths show

that mantle  $f_{O_2}$  decreases with increasing pressure<sup>13</sup>, such that carbonates or carbonatites are not stable at depths greater than ~120 km in subcratonic and asthenospheric mantle<sup>14,15</sup>. Fe<sup>2+</sup>/Fe<sup>3+</sup> equilibria in a predominantly magnesian mantle are inefficient in buffering  $f_{O_2}$  because slight changes in the Fe<sup>2+</sup>/Fe<sup>3+</sup> ratio have a strong impact on mantle  $f_{O_2}$  (ref. 16). Oxygen fugacity in the upper mantle is thus affected by processes such as partial melting, mantle metasomatism, and recycling of oxidized material by subduction. Small amounts of admixed oxidized component may raise the  $f_{O_2}$  of a limited mantle domain such that carbonates become stable and  $f_{O_2}$  is controlled by equilibria like enstatite + magnesite = olivine + graphite/diamond (EMOG/EMOD)<sup>17</sup>. The onset of carbonatite melting depends then solely on the solidus of carbonated peridotite.

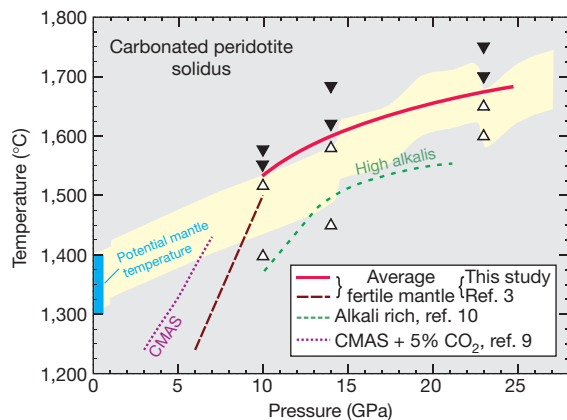
This situation changes at higher pressures: thermodynamic calculations<sup>4</sup> and experiments<sup>5,6,18</sup> demonstrate that  $f_{O_2}$  decreases with increasing pressure such that (Fe,Ni)-metal probably becomes stable at ~250 km depth<sup>6,14</sup> and in all mantle regions below<sup>5</sup>. At metal saturation depths,  $f_{O_2}$  in the Earth's mantle becomes narrowly constrained. Given equilibrium between mantle phases with molar Mg/(Mg+Fe<sup>2+</sup>) (=  $X_{Mg}$ ) ≈ 0.90 and (Fe,Ni) metal,  $f_{O_2}$  can only vary from values around the iron–wüstite (IW) equilibrium where the metal would be Ni-rich, to about 1.5 logarithmic units below IW where the metal would be almost pure iron. Because these equilibria have a considerable buffer capacity and the mantle represents an almost infinite reservoir, the mantle is capable of imposing its ambient  $f_{O_2}$  on any additional redox sensitive component, such as carbonates, carbonatites or carbon–hydrogen–oxygen fluids. Redox state differences between such fluids and ambient peridotite may trigger local hydrous redox melting<sup>19</sup> at shallow mantle depths when fluid speciation changes from reduced methane-rich to oxidized water-rich.

At larger mantle depths, small amounts of H<sub>2</sub>O can be incorporated into nominally anhydrous minerals<sup>20,21</sup> and CO<sub>2</sub> dominates the lowering of melting temperatures through volatiles<sup>22</sup>. To explore whether redox equilibria involving recycled carbonates trigger the deepest melting in the Earth's mantle, we investigated a carbonated fertile mantle at pressures (10–23 GPa), temperatures (1,400–1,900 °C) and redox conditions relevant for the mid-upper to lower mantle using high-pressure multianvil devices (Methods).

The first set of experiments constrains the solidus of a carbonated fertile peridotite at 10, 14 and 23 GPa to temperatures of 1,535, 1,600 and 1,675 °C (Fig. 1). Melting temperatures determined previously vary considerably, mainly because of variations in the bulk alkali contents and the use of synthetic analogues<sup>3,9–12</sup>. Our 10 GPa solidus temperature corresponds well to the ~1,500 °C reported previously for 10 GPa in a study on natural carbonated peridotite<sup>3</sup>, and thus provides a direct link between our data set and lower-pressure data. The solidus from 10 to 23 GPa proceeds parallel to that of alkali-enriched peridotite at 10–20 GPa (ref. 10), although absolute temperatures obtained with our composition are 100–150 °C higher (Fig. 1). This difference results from our considerably lower, peridotite-like, alkali content and a therefore lower alkali/CO<sub>2</sub> ratio (0.06 this study; 0.18 in ref. 10), which is potentially a key variable controlling solidus temperatures<sup>3</sup>.

<sup>1</sup>Institut für Geochemie und Petrologie, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland.





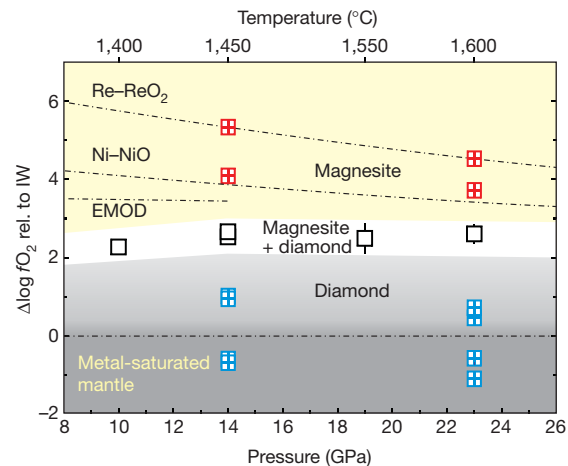
**Figure 1 | The solidus of carbonated peridotite.** Solid red line, solidus as bracketed by our experiments. Subsolidus runs, open triangles; runs containing carbonatite melt, filled inverted triangles. Run pressure uncertainties are smaller than symbol size; for temperature uncertainties see Methods. Dashed lines, solidi of natural peridotite; dotted line, solidus of a synthetic  $\text{CaO-MgO-Al}_2\text{O}_3\text{-SiO}_2$  (CMAS) +  $\text{CO}_2$  analogue system<sup>9</sup>. Brown long-dashed line, natural peridotite + 2.5 wt%  $\text{CO}_2$  (ref. 3); green short-dashed line, natural peridotite + 5 wt%  $\text{CO}_2$  enriched with 0.5 wt%  $\text{Na}_2\text{O}$  + 0.4 wt%  $\text{K}_2\text{O}$  (ref. 10). Solidi temperatures decrease with increasing bulk alkali/ $\text{CO}_2$  ratio<sup>3</sup>, although K appears to have a much stronger effect than Na (compare refs 3 and 10 with similar alkali/ $\text{CO}_2$ ). Blue bar, range of potential mantle temperatures at 1 bar extrapolated to depth (yellow corridor) assuming adiabatic behaviour in a convecting mantle<sup>32</sup>. The solidus of carbonated peridotite approaches the geotherm at  $\sim 10$  GPa (ref. 3). As at higher pressures the solidus continues within the range of adiabatic temperatures, we suggest that carbonatitic melts may form at ambient mantle temperatures over a large pressure interval if carbonate is stable with respect to  $f_{\text{O}_2}$ .

The steep positive slope of the solidus in a pressure–temperature ( $P$ – $T$ ) plot— $55^\circ\text{C}$  per GPa up to 10 GPa (ref. 3)—changes between 10 and 15 GPa to a rather weak temperature dependence of  $12^\circ\text{C}$  per GPa at  $>15$  GPa. We attribute this change to an increase in activity with pressure of mainly  $\text{Na}_2\text{O}$  but also  $\text{CaO}$ , caused by the continuous decrease in modal abundance of clinopyroxene. Above 15 GPa, pyroxene completely dissolves in majorite–garnet solid solution<sup>23</sup>. Sodium is relatively incompatible in the garnet structure and partitions strongly into a carbonatite liquid, thus causing relatively low melting temperatures. The solidus of carbonated peridotite approaches the mantle geotherm at about 10 GPa and from then on remains close to ambient adiabatic mantle temperatures (Fig. 1). Carbonatite melting in the deep Earth's mantle does therefore not require anomalously high temperatures and carbonatites may be produced over large depth intervals.

In two further series of experiments, we determined carbon speciation as a function of  $f_{\text{O}_2}$  for slightly subadiabatic mantle temperatures. For this, we first equilibrated the carbonated peridotite with Fe–FeO, Ni–NiO, (Ni,Au)–NiO and Re– $\text{ReO}_2$  solid state metal–metal oxide buffers at 14 GPa,  $1,450^\circ\text{C}$  and 23 GPa,  $1,600^\circ\text{C}$ . Fe, Ni, Ni–Au and Re metal were used as capsule materials, and the respective oxide was either contained in the starting material, or (in the case of  $\text{ReO}_2$ ) added to the charge. Oxygen fugacities imposed by these buffers range from  $-1$  to  $+5$  log units relative to IW (Methods). Experiments under oxidizing conditions (IW  $+4$  to  $+5$ ) in Re and Au–Ni capsules yield almost pure magnesite ( $X_{\text{Mg}} = 0.96$ ), whereas runs in Fe and Ni capsules reveal that carbonate is not stable at  $f_{\text{O}_2} < \text{IW} + 1.2$  at either 14 or 23 GPa (Fig. 2). Instead, carbonate is reduced to micrometre-sized diamonds, identified in backscatter electron images and measured using energy-dispersive spectroscopy. Elevated FeO and NiO contents and fractions in the minerals provide further evidence for carbonate reduction via



where  $(\text{Fe,Ni,Mg})\text{O}$  is a compound in olivine, garnet and perovskite or is ferropericlase.



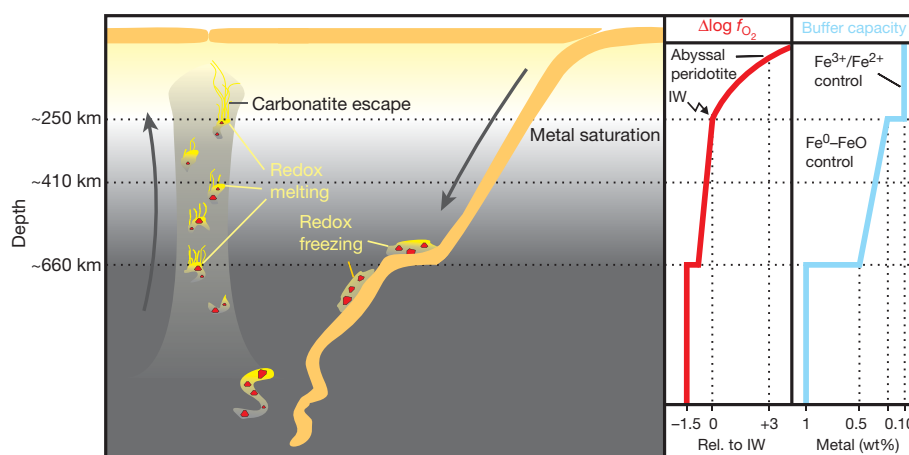
**Figure 2 | Carbon speciation in natural mantle as a function of pressure, temperature and  $f_{\text{O}_2}$ .** Plotted are calculated  $\log f_{\text{O}_2}$  values relative to the IW reference (Methods) for magnesite-bearing runs (red symbols), diamond-bearing runs (blue symbols) and runs containing diamond + magnesite using Ir as redox sensor (black symbols). Coloured areas mark the stability fields of magnesite (yellow), magnesite+diamond (white) and diamond (light grey). Run durations were at least 24 h to ensure redox equilibrium between the phases (Supplementary Table 1). The stability field of magnesite remains at  $\geq 2$  log units above IW up to 23 GPa; no strong pressure dependence is evident from our data. At the low  $f_{\text{O}_2}$  conditions estimated for the Earth's mantle which lead to metal saturation at pressures higher than  $\sim 8$  GPa (corresponding to  $\sim 250$  km depths; dark grey area), carbonates and carbonatites are unstable and will be reduced to diamond according to reaction (1). Relative positions of IW and Ni–NiO buffers calculated after ref. 33; for relative position of Re– $\text{ReO}_2$ , see Methods. The EMOD buffer dominating in the uppermost mantle<sup>17</sup> (equation given in the text) is limited to  $\sim 14$  GPa by the stability of olivine and pyroxene. Run pressure uncertainties are within the size of the symbols, as are propagated  $2\sigma$  errors for  $f_{\text{O}_2}$ , unless error bars are reported. Errors include analytical uncertainties and  $2\sigma$  standard deviations from averages for runs where  $f_{\text{O}_2}$  was calculated for multiple phases separately, but do not include uncertainties of the activity models which are difficult to quantify. Because of this, we suggest that an additional uncertainty for all  $f_{\text{O}_2}$  values reported is  $\pm 0.3$  log units.

Additionally, we monitored  $f_{\text{O}_2}$  in experiments saturated in both diamond and carbonate at 10–23 GPa by adding 5 wt% iridium metal as a redox sensor<sup>24</sup>. The Fe content in Ir, together with the FeO content in silicates or oxides, allows the calculation of  $f_{\text{O}_2}$  conditions during the experiment (Methods). The carbonated peridotite was saturated with elemental carbon through the use of graphite capsules that transformed to diamond or, for noble metal capsules, through addition of 10 wt% C. We also added 3 wt% Fe<sup>0</sup> to enhance the mass reacted through equilibrium (1). These redox sensing experiments yield  $f_{\text{O}_2}$  values slightly lower than calculated for the EMOD equilibrium at 10 and 14 GPa and then remain rather constant at 2.3–2.7 log units above IW up to 23 GPa. This is consistent with the  $f_{\text{O}_2}$  corridor between IW  $+1.2$  to  $+4$ , as defined by the externally buffered experiments, in which the coexistence of diamond+diamond would be possible (Fig. 2).

Our results imply that at an ambient, fertile mantle  $f_{\text{O}_2}$  around IW  $-1.5$  (ref. 14), carbonate that is remixed into the mantle at  $>250$  km depth is unstable and will be reduced to diamond. Typical carbon concentrations of 20–250 p.p.m. C for sub-ridge mantle (ref. 25 and references therein) would inevitably dissolve in the metal phase present at these depths or form discrete iron carbides, like  $\text{Fe}_3\text{C}$  and  $\text{Fe}_7\text{C}_3$ , depending on Fe–C ratio and  $P$ – $T$  conditions<sup>25,26</sup>. Thus, carbonate related melting is unlikely to occur in Earth's lower mantle, the transition zone and the lowermost upper mantle as long as sub-ridge carbon concentrations prevail (Fig. 2). Although the mantle ceases to be metal saturated in shallower parts of the upper mantle ( $<250$  km, ref. 6), the average  $f_{\text{O}_2}$  up to 100–150 km depth presumably remains too low to sustain carbonates or carbonatites as equilibrium phases at sub-ridge carbon concentrations<sup>15</sup>.

However, the addition of subducted carbonate to average mantle fundamentally changes this behaviour. Relative buffer capacity changes in the Earth's mantle induce immobilization of carbonatite melts through 'redox freezing', that is, reduction of carbonatites to diamond, as well as remobilization of carbon through redox melting. Redox melting transforms diamond to carbonatite melts, which potentially control the onset of ultra-deep melting. Starting from a subducting, locally carbonated, relatively oxidized mafic to ultramafic lithosphere, our experiments demonstrate that carbonatite melts will be generated in such lithosphere on thermal relaxation (Fig. 1). This may occur when the lithosphere deflects into the transition zone above the 660-km discontinuity or when stagnating in the lower mantle. On a local scale, oxidized carbonatite melt migrating into the mantle will consume metal to first form iron carbide in an intermittent stage, and then further oxidize the Fe and Ni contained in the carbide to leave a mantle domain that contains all iron as  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  in silicates and ferropicls and all carbon as diamond. Owing to its low viscosity and high wetting properties<sup>27,28</sup>, any excess carbonatite not consumed by redox reactions would percolate upwards along grain boundaries and exhaust further (Fe,Ni)-metal and carbide until complete redox freezing—that is, immobilization due to reduction of  $\text{CO}_2$  to  $\text{C}^0$ —is achieved. This presumably very efficient process will eventually exhaust all buffering metal and carbide through precipitation of diamond, and result in a metal-free mantle domain where diamond coexists with  $\text{Fe}^{3+}$ -bearing garnet or perovskite (Fig. 3).

At the boundary of such domains, where the supply of carbonatite melt does not exceed the redox capacity of Fe,Ni-metal, an iron carbide rim is expected to form. The redox capacity of  $\text{Fe}^{3+}$  in such mantle domains would be exactly equivalent or slightly superior (due to the  $\text{Fe}^{3+}$  present before disproportionation of  $\text{Fe}^{2+}$ ) to that necessary to re-convert all diamond to  $\text{CO}_2$ . Similarly, the maximum increase in C content of mantle domains metasomatized by carbonatites derived from the subducting lithosphere is restricted to  $\sim 1,000$  p.p.m., equivalent to the amount of carbonatite that may be immobilized by the maximum Fe-metal content of 1 wt% expected in the lower mantle<sup>5</sup>.



**Figure 3 | Carbonatitic redox freezing and redox melting caused by redox capacity changes in Earth's mantle.** Main panel, cartoon illustrating a possible sequence of redox freezing and redox melting events driven by oxidation state contrasts between subducted lithosphere and ambient asthenospheric mantle. Right, potential mantle  $f_{\text{O}_2}$  (red line) and redox buffer capacity (blue line) as function of depth.  $f_{\text{O}_2}$  decreases in the subcratonic upper mantle<sup>13</sup> to reach  $\sim \text{IW}$  and metal saturation at  $\sim 250$  km depth<sup>6,14</sup>. Change in redox control from  $\text{Fe}^{2+}/\text{Fe}^{3+}$  in the upper mantle to  $\text{Fe}^0\text{--FeO}$  provides a considerable gain in  $f_{\text{O}_2}$  buffer capacity. At  $>250$  km depth, the mantle and all redox sensitive components are bound to  $f_{\text{O}_2}$  levels  $< \text{IW}$  unless all metal is exhausted. Majoritic garnet in equilibrium with  $\text{Fe}^0$  incorporates more  $\text{Fe}^{3+}$  with increasing pressure<sup>6</sup>; adequate charge balance results from reduction of  $\text{FeO}$  to  $\text{Fe}^0$  which increases the  $\text{Fe}^0$  metal abundance. An estimate yields 0.5 wt% metal at the base of the transition zone, assuming all Ni is present as  $\text{Ni}^0$  (1,800 p.p.m.)

As the metal content is expected to decrease with decreasing pressure (Fig. 3), the corresponding maximum C content of metasomatized domains would also be lower for domains formed at shallower depth.

The inverse process, redox melting, would occur when such mantle heterogeneities are entrained by upwelling mantle and cross the 660-km discontinuity and transition zone. Destabilization of  $\text{Fe}^{3+}$ -rich perovskite without an adequate amount of  $\text{Fe}^0$  for comproportionation (the opposite process to disproportionation) inevitably leads to a sudden increase in  $\text{Fe}^{3+}$  activity at the 660-km discontinuity and thus to the re-oxidation of diamond to  $\text{CO}_2$ . This redox reaction would directly result in carbonatite melts within upwelling, previously melt-free but diamond-bearing mantle. Carbonatitic redox melting might also explain the virtual absence of pyrolitic mineral inclusions in diamonds originating from sublithospheric depths ( $>200\text{--}300$  km)<sup>29</sup>. Pyrolitic pristine mantle itself does not contain sufficient C to form diamonds<sup>25</sup>, and thus diamonds are expected to form dominantly through the above redox freezing process wherever carbonatites percolate. The inverse process—that is, redox melting which destroys most of these diamonds through oxidation—then leads to the generation of melts that carry remnants of such deep mantle domains to the Earth's surface. Average mantle is expected to contain  $\sim 1$  wt%  $\text{Fe}^0$  formed from  $\text{Fe}^{2+}$  disproportionation in the lower mantle<sup>5</sup>; such metal fractions are stable and would not segregate to Earth's core<sup>30</sup>. For reaction (1), the redox capacity of 1 wt%  $\text{Fe}^0$  is equivalent to 0.8 wt% magnesite. As argued above, the properties of carbonatite melt result in a self-regulating mechanism, where infiltrating carbonatite melt oxidizes all  $\text{Fe}^0$ , leaving behind a diamond-bearing mantle domain with exactly the same redox-capacity (that is, that of 0.8 wt% magnesite). In the reverse process, it can thus be expected that about 1 wt% of carbonatite melt forms in such upwelling mantle domains. Although upwelling occurs at speeds comparable to plate tectonic movements (that is,  $1\text{--}10$  cm  $\text{yr}^{-1}$ ), 1% low viscosity melt in the mantle matrix rises with speeds of at least  $10\text{--}100$  m  $\text{yr}^{-1}$  (ref. 31). Consequently, carbonatite flow will tend to escape from the upwelling mantle matrix but will suffer redox freezing as long as progressing carbonatite melts encounter

plus  $\sim 3,200$  p.p.m.  $\text{Fe}^0$  calculated from mass balance (mantle: 8 wt%  $\text{FeO}^{\text{total}}$ ; 40 vol% majoritic garnet: 5.5 wt%  $\text{FeO}^{\text{total}}$ ,  $\text{Fe}^{3+}/\Sigma\text{Fe} = 0.33$  (ref. 6); 60 vol% ringwoodite: 9.7 wt%  $\text{FeO}^{\text{total}}$ ,  $\text{Fe}^{3+}/\Sigma\text{Fe} = 0.02$  (ref. 14)). Transition to the lower mantle is accompanied by a gain in buffer capacity because MgSi-perovskite incorporates more  $\text{Fe}^{3+}$  at metal saturation  $f_{\text{O}_2}$  than majoritic garnet, resulting in  $\sim 1.0$  wt% metal<sup>5</sup>. Remixing of subducted carbonated lithosphere into the mantle at these depths leads to redox freezing and diamond-enriched mantle domains. Domains incorporated into upwelling mantle will experience redox melting, most prominently at levels where the capacity of the mantle phases to incorporate  $\text{Fe}^{3+}$  changes drastically, that is, at  $\sim 660$  km and  $\sim 250$  km depth. Melts produced from C-enriched domains cannot escape through the surrounding metal-bearing average mantle, but will freeze until redox control shifts from metal saturation back to bulk  $\text{Fe}^{2+}/\text{Fe}^{3+}$  controlled at  $\sim 250$  m depth.

metal-bearing mantle. It is only at depths <250 km when garnet becomes less majoritic,  $\text{Fe}^{3+}$  activity increases, and the mantle ceases to be metal-saturated, that this scenario changes. At this depth,  $f_{\text{O}_2}$  buffering in the mantle shifts back from being controlled by  $\text{Fe}^0/\text{Fe}^{2+}$  to being controlled by  $\text{Fe}^{2+}/\text{Fe}^{3+}$ , which leads to a dramatic decline in buffering capacity. Such mantle has little potential to hinder carbonatite melts from percolating upwards.

We expect that redox equilibria control the formation of deep carbonatite melts, and that at ~250 km depth a continuous flow of carbonatite melt originating from mantle heterogeneities may form in upwelling mantle, which is probably reflected in seismic low velocity anomalies<sup>1,2</sup> observed at these depths. If melt quantities are too low and evade direct seismological observation, their presence may be confirmed by enhanced electrical conductivity or seismic anisotropies caused by matrix reorganization<sup>3</sup>.

## METHODS SUMMARY

**Starting composition.** The starting composition for the experiments represents a fertile peridotite enriched in 5 wt%  $\text{CO}_2$ . NiO in this composition was set to 1 wt%, to ensure that the NiO content of the phases was high enough for microprobe analysis and  $f_{\text{O}_2}$  calculations. Experiments using Ir as a redox sensor were without Ni to ensure that the metal is a binary Fe–Ir alloy.

**Oxygen fugacity calculations.** We calculate  $f_{\text{O}_2}$  using Fe–FeO and Ni–NiO redox equilibria.  $f_{\text{O}_2}$  values are reported relative to iron–wüstite (IW), nickel–NiO (NNO) and rhennium– $\text{ReO}_2$  (RRO), using activity–composition relationship for (Fe,Ni) alloys and (Au,Ni) alloys (Methods).  $f_{\text{O}_2}$  values of ferropericlase (fp)-bearing experiments are calculated from

$$2\text{M} + \text{O}_2 = 2\text{MO} \quad (2)$$

with  $\text{M} = \text{Fe}$  or  $\text{Ni}$ .  $f_{\text{O}_2}$  relative to the respective metal–metal oxide (MMO) equilibrium is given by

$$\Delta \log f_{\text{O}_2} [\text{MMO}] = 2 \log a_{\text{MO}}^{\text{fp}} - 2 \log a_{\text{M}}^{\text{metal}} \quad (3)$$

where activity  $a$  is defined as molar fraction  $X$  times activity coefficient  $\gamma$ .  $f_{\text{O}_2}$  conditions of runs containing olivine, wadsleyite or ringwoodite were calculated from Fe activity in the olivine polymorph and the coexisting metal phase using binary symmetric solution models for activity corrections.

**Iridium redox sensing.** Ir metal was used as a redox sensor<sup>24</sup> to determine  $f_{\text{O}_2}$  in runs containing elemental carbon+magnesite. Margules interaction parameters were obtained by least-squares fitting of a binary asymmetric regular solution model to X-ray data for face-centred cubic (f.c.c.) Ir–Fe alloy (Methods). Activity coefficients  $\gamma$ , dependent on  $P$ ,  $T$  and  $X$  for Fe in f.c.c. Ir–Fe alloy, are given by

$$\text{RT} \ln(\gamma_{\text{Fe}}) = 2X_{\text{Fe}}X_{\text{Ir}}W_{\text{G Ir-Fe}} + X_{\text{Ir}}^2W_{\text{G Fe-Ir}} - 2G^{\text{xs}} \quad (4)$$

with  $G^{\text{xs}}$  being the excess Gibbs free energy of mixing, and  $W_{\text{G Ir-Fe}}$ ,  $W_{\text{G Fe-Ir}}$  being the Margules parameters. Correcting the activity model for  $P$  and  $T$  is described in Methods and shown in Supplementary Fig. 1.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 28 May 2010; accepted 31 January 2011.

Published online 23 March 2011.

1. Forsyth, D. W. *et al.* Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science* **280**, 1215–1218 (1998).
2. Gu, Y. J., Lerner-Lamb, A. L., Dziewonski, A. M. & Ekström, G. Deep structure and seismic anisotropy beneath the East Pacific Rise. *Earth Planet. Sci. Lett.* **232**, 259–272 (2005).
3. Dasgupta, R. & Hirschmann, M. M. Melting in the Earth's deep upper mantle caused by carbon dioxide. *Nature* **440**, 659–662 (2006).
4. Ballhaus, C. Is the upper mantle metal-saturated? *Earth Planet. Sci. Lett.* **132**, 75–86 (1995).
5. Frost, D. J. *et al.* Experimental evidence for the existence of iron-rich metal in the Earth's lower mantle. *Nature* **428**, 409–412 (2004).
6. Rohrbach, A. *et al.* Metal saturation in the upper mantle. *Nature* **449**, 456–458 (2007).

7. Allègre, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. *Nature* **323**, 123–127 (1986).
8. Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
9. Dalton, J. A. & Presnall, D. C. Carbonatitic melts along the solidus of model lherzolite in the system  $\text{CaO-MgO-Al}_2\text{O}_3\text{-SiO}_2\text{-CO}_2$  from 3 to 7 GPa. *Contrib. Mineral. Petrol.* **131**, 123–135 (1998).
10. Ghosh, S., Ohtani, E., Litasov, K. D. & Terasaki, H. Solidus of carbonated peridotite from 10 to 20 GPa and origin of magnesio-carbonatite melt in the Earth's deep mantle. *Chem. Geol.* **262**, 17–28 (2009).
11. Brey, G. P., Bulatov, V. K., Gurnis, A. V. & Lahaye, Y. Experimental melting of carbonated peridotite at 6–10 GPa. *J. Petrol.* **49**, 797–821 (2008).
12. Litasov, K. D. & Ohtani, E. Solidus and phase relations of carbonated peridotite in the system  $\text{CaO-Al}_2\text{O}_3\text{-MgO-SiO}_2\text{-Na}_2\text{O-CO}_2$  to the lower mantle depths. *Phys. Earth Planet. Inter.* **177**, 46–58 (2009).
13. Woodland, A. B. & Koch, M. Variation in oxygen fugacity with depth in the upper mantle beneath the Kaapvaal craton, South Africa. *Earth Planet. Sci. Lett.* **214**, 295–310 (2003).
14. Frost, D. J. & McCammon, C. A. The redox state of the Earth's mantle. *Annu. Rev. Earth Planet. Sci.* **36**, 389–420 (2008).
15. Stagno, V. & Frost, D. J. Carbon speciation in the asthenosphere: experimental measurements of the redox conditions at which carbonate-bearing melts coexist with graphite or diamond in peridotite assemblages. *Earth Planet. Sci. Lett.* **300**, 72–84 (2010).
16. Gudmundsson, G. & Wood, B. J. Experimental tests of garnet peridotite oxygen barometry. *Contrib. Mineral. Petrol.* **119**, 56–67 (1995).
17. Eggler, D. H. & Baker, D. R. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.) 237–250 (Centre for Academic Publishing, 1982).
18. O'Neill, H. St C. *et al.* in *Evolution of the Earth and Planets* (eds Takahashi, E., Jeanloz, R. & Rubie, D. C.) 73–88 (Geophys. Monogr. 74, Int. Union Geol. Geophys./Am. Geophys. Union, 1993).
19. Taylor, W. R. & Green, D. H. in *Magmatic Processes and Physicochemical Principles* (ed. Mysen, B. O.) 121–138 (Spec. Publ. No. 1, Geochem. Soc. USA, 1987).
20. Kohn, S. C. Solubility of  $\text{H}_2\text{O}$  in nominally anhydrous mantle minerals using  $^1\text{H}$  MAS NMR. *Am. Mineral.* **81**, 1523–1526 (1996).
21. Keppler, H. & Bolfan-Casanova, N. in *Water in Nominally Anhydrous Minerals* (eds Keppler, H. & Smyth, J. R.) 193–230 (Reviews in Mineralogy and Geochemistry 62, Min. Soc. Am., 2006).
22. Luth, R. W. in *Mantle Petrology: Field Observations and High Pressure Experimentation: A Tribute to Francis R. (Joe) Boyd* (eds Fei, Y., Bertka, C. M. & Mysen, B. O.) 297–316 (Geochemical Society, 1999).
23. Ringwood, A. E. The pyroxene–garnet transformation in the Earth's mantle. *Earth Planet. Sci. Lett.* **2**, 255–263 (1967).
24. Woodland, A. B. & O'Neill, H. St C. Thermodynamic data for Fe-bearing phases obtained using noble metal alloys as redox sensors. *Geochim. Cosmochim. Acta* **61**, 4359–4366 (1997).
25. Dasgupta, R. & Hirschmann, M. M. The deep carbon cycle and melting in Earth's interior. *Earth Planet. Sci. Lett.* **298**, 1–13 (2010).
26. Lord, O. T., Walter, M. J., Dasgupta, R., Walker, D. & Clark, S. M. Melting in the Fe–C system to 70 GPa. *Earth Planet. Sci. Lett.* **284**, 157–167 (2009).
27. Minarik, W. G. & Watson, E. B. Interconnectivity of carbonate melt at low melt fraction. *Earth Planet. Sci. Lett.* **133**, 423–437 (1995).
28. Hunter, R. H. & McKenzie, D. The equilibrium geometry of carbonate melts in rocks of mantle composition. *Earth Planet. Sci. Lett.* **92**, 347–356 (1989).
29. Stachel, T., Brey, G. P. & Harris, J. W. Inclusions in sublithospheric diamonds: glimpses of deep earth. *Elements* **1**, 73–78 (2005).
30. Bagdassarov, N., Solferino, G., Golabek, G. J. & Schmidt, M. W. Centrifuge assisted percolation of Fe–S melts in partially molten peridotite: time constraints for planetary core formation. *Earth Planet. Sci. Lett.* **288**, 84–95 (2009).
31. Connolly, J. A. D., Schmidt, M. W., Solferino, G. & Bagdassarov, N. Permeability of asthenospheric mantle and melt extraction rates at mid-ocean ridges. *Nature* **462**, 209–212 (2009).
32. Stixrude, L. & Lithgow-Bertelloni, C. Influence of phase transformations on lateral heterogeneity and dynamics in Earth's mantle. *Earth Planet. Sci. Lett.* **263**, 45–55 (2007).
33. Campbell, A. J. *et al.* High pressure effects on the iron–iron oxide and nickel–nickel oxide oxygen fugacity buffers. *Earth Planet. Sci. Lett.* **286**, 556–564 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Discussions with P. Ulmer, U. Mann and C. Ballhaus stimulated this work. Comments and suggestions from T. Stachel improved the manuscript. Financial support by Swiss National Science Foundation (SNSF) grant 2-777-86-06 is acknowledged.

**Author Contributions** M.W.S. and A.R. designed this project and M.W.S. obtained its funding. Experiments, analytical work and  $f_{\text{O}_2}$  calculations were done by A.R.; both authors contributed equally to all other parts.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to A.R. ([arno.rohrbach@gmail.com](mailto:arno.rohrbach@gmail.com)).



## METHODS

**Starting material.** The starting compositions for multianvil experiments used in this study represent a fertile peridotite composition<sup>34</sup> enriched in 5 wt% CO<sub>2</sub> (Supplementary Table 2). Starting composition pyr5 was used to determine the solidus temperature of carbonated peridotite and for runs where the oxygen fugacity of the charge was controlled by the Fe, Ni, and (Au,Ni) capsule materials. Both NiO and CoO contents of pyr5 were set to 1 wt%, to ensure that NiO contents of the phases were high enough for microprobe analysis and  $f_{O_2}$  calculation from these data. Experiments using Ir metal as redox sensor were performed with the simplified pyr5 starting material to ensure that the metal phase in the charge is a binary Fe–Ir alloy.

Starting composition pyr5 was prepared from reagent grade oxides SiO<sub>2</sub>, TiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, Cr<sub>2</sub>O<sub>3</sub>, MgO, fired overnight at 1,000 °C before mixing and grinding under acetone in an agate mortar. Part of the CaO inventory was added to the oxide mix as reagent grade CaCO<sub>3</sub> (dried at 220 °C for 5 h), mixed and decarbonized by stepwise heating to 1,000 °C. Afterwards, iron was added as synthetic fayalite, NiO and CoO as oxides, all dried at 220 °C for 5 h. CO<sub>2</sub> was added as a mixture of reagent grade CaCO<sub>3</sub>, MgCO<sub>3</sub> and Na<sub>2</sub>CO<sub>3</sub> dried at 220 °C for 5 h in 61/34/5 wt ratio. All components were mixed and ground again thoroughly under acetone in an agate mortar.

Starting composition pyr5 was prepared from reagent grade SiO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, MgO (all fired at 1,000 °C for 5 h) synthetic fayalite, synthetic wollastonite, natural magnesite (kindly provided by P. Ulmer), and reagent grade Na<sub>2</sub>CO<sub>3</sub> (all dried at 220 °C for 5 h).

To ensure nominally anhydrous conditions, the starting materials were stored in an desiccator together with silica gel; synthetic rock powders were dried for 2 h at 220 °C before pressing it into capsules and the filled capsule was dried again for 1 h at 220 °C before final welding.

**Multianvil experiments.** High pressure experiments (Supplementary Table 1) were performed in 600 t and 1,000 t Walker type multianvil devices. We used 32-mm tungsten carbide cubes as second stage anvils and natural pyrophyllite gaskets as pressure transmitting media. We used an 18/11 assembly for runs at 10 GPa (18 mm octahedron edge length, 11 mm truncated edge length), a 14/8 assembly for 14 GPa and a 10/3.5 assembly for 19 and 23 GPa. All assemblies consist of a Cr<sub>2</sub>O<sub>3</sub> doped MgO octahedron with LaCrO<sub>3</sub> heater (stepped in case of 18/11 and 14/8), ZrO<sub>2</sub> thermal insulation, MgO spacers, and Mo metal plugs to ensure electrical contact between assembly and WC cubes. Pressure was computer controlled during the entire run; temperature was measured with B- or C-type thermocouples and controlled with a Eurotherm controller to about ±5 °C. Experiments were quenched (by turning off the power supply) at a rate of ~800 °C per second. The total temperature precision including temperature gradients within the capsule might be ±15 °C in 18/11 and 14/8 experiments. Temperature gradients within the 10/3.5 assembly might be 30–50 °C over the capsule length of ~1.2 mm. Details of pressure calibration, design, and temperature gradients of the 10/3.5 assembly are given in ref. 35.

**Encapsulation.** Runs to determine the carbonated peridotite solidus were performed in welded Au<sub>80</sub>Pd<sub>20</sub> capsules. No attempt was made to control  $f_{O_2}$  during these runs. The use of conventional Pt–C double capsules appeared to be problematic as the space is very limited and recovering the sample from the graphite capsule that transformed to diamond during the runs was rather challenging. Mass balance calculations and measured Fe contents in the capsule material show that iron loss from the charge to the Au<sub>80</sub>Pd<sub>20</sub> capsule is negligible (see also  $X_{Mg}$  values in Supplementary Table 2).

Runs using Ir metal as redox sensor were performed in welded Au<sub>80</sub>Pd<sub>20</sub> capsules. We added ~10 wt% of graphite to the charges to obtain the  $f_{O_2}$  buffering through coexisting C<sup>0</sup> and carbonate in a peridotite matrix. We added ~5 wt% fine grained Ir metal and 3–5 wt% of metallic Fe to ensure redox equilibrium via carbonate reduction according to reaction (1).

Slightly lower  $X_{Mg}$  of all phases in  $f_{O_2}$  monitoring runs (Supplementary Table 2) are indicative for reaction (1) and therefore for the attainment of redox equilibrium in these runs. Both diamond and magnesite were present in experimental charges in direct contact with each other as verified by backscattered imaging, EDX and EPMA analysis. This encapsulation method yields results that agree within error to those from experiments in C capsules surrounded by Re metal foil (compare  $f_{O_2}$  of runs 39 and 47; Supplementary Table 5) but sample recovery and polishing was facilitated considerably, especially for runs at 23 GPa.

Fe<sup>0</sup> and Ni<sup>0</sup> capsules were not welded but cold sealed at run pressure with a conical cap. We added ~3 wt% of the respective metal to the charge to facilitate attainment of equilibrium.

(Au,Ni) alloy capsules were prepared as welded Au outer capsule and Ni foil inner capsule which formed an homogeneous (Au,Ni) alloy at run pressures and temperatures. In runs 22 and 29b we added ~10 wt% NiO in order to raise the NiO content of the phases and accordingly relative  $f_{O_2}$ . This NiO-enriched compositions crystallized a NiO-rich periclase instead of ferropericlase at 23 GPa (run 29b) and an olivine enriched in Ni<sub>2</sub>SiO<sub>4</sub>-component at 14 GPa (run 22, see Supplementary Table 2).

**Analytical methods.** Recovered sample capsules (in case of 10/3.5 runs, complete octahedra) were mounted in an epoxy resin, ground down with sand paper and polished with polycrystalline diamond paste. Runs to determine the solidus were polished under dry conditions using synthetic polishing paper.

Runs were analysed with a Jeol JXA8200 microprobe using minerals and synthetic oxides as standards. Results are given in Supplementary Table 2. Silicates and oxides were measured at 15 kV and 20 nA with a focused beam, counting times were 20 s on the peak and 10 s on the background. Perovskites and carbonates were measured with an electron beam of 5 µm diameter at 10 nA to minimize radiation damage. Note, however that CaSi-perovskite was still unstable under the beam, which results in low totals. Additionally, CaSi-perovskite often shows a blurred exsolution texture that is difficult to analyse without obtaining mix signals. The majority of the experiments, especially under subsolidus conditions (despite run durations up to 4 days) showed average grain sizes between 3 and 10 µm and did not allow a broad beam. In experiments AR13 and AR18, some phases were too small to be measured by electron microprobe and were identified using backscattered electron images and EDX (Jeol JSM 6390LA scanning electron microscope).

Metals were measured with a focused beam at 15 kV and 20 nA, counting times were 30 s on the peak and 10 s on the background. We used Fe, Ir, Ni and Co metal as standards for quantification to reduce matrix effects. To minimize the excitation volume of the sample, to be able to measure small Ir–Fe alloy grains, we measured Ir at 15 kV and 20 nA using its low excitation energy Mα peak.

**Calculation of oxygen fugacities.** We calculated  $f_{O_2}$  of the experimental charges using Fe–FeO and Ni–NiO redox equilibria.  $f_{O_2}$  values are reported relative to iron-wüstite (IW) and nickel–NiO (NNO) buffers. The relative positions of IW and NNO at *P* and *T* were calculated after ref. 33, the relative position of the Re–ReO<sub>2</sub> (RRO) buffer was estimated from ref. 36. The activity–composition relationship for (Fe,Ni) alloys were taken from ref. 37, activity data for (Au,Ni) alloys were adopted from ref. 38.

**Periclase-bearing experiments.** Oxygen fugacities of ferropericlase (fp)-bearing experiments (19 and 23 GPa) were calculated using equilibria of the form



with *M* = Fe or Ni.  $f_{O_2}$  relative to the respective metal–metal oxide (MMO) equilibrium reads:

$$\Delta \log f_{O_2} [MMO] = 2 \log a_{MO}^{fp} - 2 \log a_M^{metal} \quad (6)$$

where activity (*a*) is defined as molar fraction (*X*) times activity coefficient ( $\gamma$ ). The molar fraction for metals and phases, calculated activity coefficients and calculated  $f_{O_2}$  values of the experiments are summarized in Supplementary Tables 3–5.  $a_{FeO}^{fp}$  was determined using a binary regular solution model:

$$RT \ln(\gamma_{FeO}^{fp}) = (11,000 + 0.011P)(1 - X_{FeO}^{fp})^2 \quad (7)$$

for FeO, where *P* is in bar, *T* in K, and *R* is the gas constant<sup>39</sup>. We note that ferric iron content of ferropericlase was not measured and microprobe FeO<sup>total</sup> data were used to calculate  $X_{Fe}$ . The  $Fe^{3+}/\Sigma Fe$  for ferropericlase in systems with  $X_{Mg} = 0.9$  however is generally <0.05 (refs 40–42), so calculated  $f_{O_2}$  values might therefore be overestimated by a maximum of 0.1 log units.

$a_{NiO}^{fp}$  was calculated using a ternary regular solution model for NiO:

$$RT \ln(\gamma_{NiO}^{fp}) = W_{Ni-Mg}(X_{Mg})^2 + W_{Ni-Mg}(X_{Fe})^2 + (W_{Ni-Mg} + W_{Ni-Fe} - W_{Mg-Fe})X_{Mg} \quad (8)$$

where *T* is in K, and *R* is the gas constant<sup>43,44</sup>.

**Olivine-, wadsleyite- or ringwoodite-bearing experiments.**  $f_{O_2}$  conditions relative to the quartz–fayalite–iron (QFI) equilibrium of runs containing olivine, wadsleyite or ringwoodite (10, 14 and 19 GPa) were calculated using the following expressions:



$$\Delta \log f_{O_2} [QFI] = 2 \log a_{Fe}^{ol/wad/rw} - 2 \log a_{Fe}^{metal} - \log a_{SiO_2} \quad (10)$$

Activity coefficients were calculated using a binary symmetric solution model where:

$$RT \ln(\gamma_{Fe}^{ol}) = W_{Fe-Mg}^{ol}(1 - X_{Fe}^{ol})^2 \quad (11)$$

Margules parameters including  $W_V$  to correct the olivine interaction parameter for pressure were taken from ref. 45 (Supplementary Table 6).

NiO-bearing olivines show a slight negative deviation from ideal mixing, even at the high temperatures relevant for our experiments<sup>46,47</sup>. Owing to this we set  $\gamma_{\text{NiO}}^{\text{ol}} = 0.9$ ; but the effect on calculated  $f_{\text{O}_2}$  is minor, that is, in the range  $\pm 0.05$  log units. Silica activities ( $a_{\text{SiO}_2}$ ) in experiments containing olivine and clinoenstatite were calculated from the equilibrium:



where

$$\log a_{\text{SiO}_2} = \log a_{\text{en}}^{\text{opx}} - \log a_{\text{fo}}^{\text{ol}} - \log K \quad (13)$$

Enstatite activity in pyroxene was calculated after ref. 48 and forsterite activity in olivine after ref. 45 as described above for fayalite.  $\log K$  of reaction (12) was calculated using thermodynamic data<sup>49</sup>.  $f_{\text{O}_2}$  values relative to IW are obtained by adding the difference in  $f_{\text{O}_2}$  between QFI and IW equilibria at run  $P$  and  $T$  to  $\Delta \log f_{\text{O}_2}$  [QFI]. **Activity-composition relation in the system Fe–Ir at high  $P$  and  $T$ .** Iridium metal was used as a redox sensor<sup>24</sup> to monitor  $f_{\text{O}_2}$  in runs that contain both elemental carbon and magnesite between 10 and 23 GPa. Margules interaction parameters ( $W$ ) dependent on  $P$ ,  $T$  and  $X$  are calculated according to the expression

$$W_G = W_{H \text{ 1bar}} - TW_S + (P - 1)W_V \quad (14)$$

where  $T$  is in K and  $P$  in bar<sup>50</sup>.

$W_{H \text{ 1bar}}$  was obtained by least-squares fitting of a binary asymmetric regular solution model<sup>51</sup> to the non-temperature-dependent part of the excess Gibbs energy ( $G^{\text{xs}}$ ) for face centred cubic (f.c.c.) Ir–Fe alloy<sup>52</sup>.  $W_S$  is  $-5 \text{ J mol}^{-1} \text{ K}^{-1}$  according to the temperature dependence of the excess Gibbs energy expression for f.c.c. Ir–Fe alloy<sup>52</sup>.  $W_V$  is estimated by least-squares fitting of a binary asymmetric regular solution model to excess volumes of the Fe–Ir alloy derived from X-ray data<sup>52</sup>. Activity coefficients  $\gamma$  dependent on  $P$  and  $T$ , and  $X$  values for Fe in f.c.c. Ir–Fe alloy result from a binary asymmetric regular solution model<sup>51</sup>:

$$RT \ln(\gamma_{\text{Fe}}^{\text{metal}}) = 2X_{\text{Fe}}X_{\text{Ir}}W_{G \text{ Ir–Fe}} + (X_{\text{Ir}})^2W_{G \text{ Fe–Ir}} - 2G^{\text{xs}} \quad (15)$$

where

$$G^{\text{xs}} = X_{\text{Fe}}X_{\text{Ir}}(X_{\text{Ir}}W_{G \text{ Fe–Ir}} + X_{\text{Fe}}W_{G \text{ Ir–Fe}}) \quad (16)$$

The effect of correcting the activity model for  $P$  and  $T$  is shown in Supplementary Fig. 1.

34. Palme, H. & O'Neill, H. St C. in *The Mantle and The Core* Vol 2. (ed. Carlson, R. W.) 1–38 (Elsevier-Pergamon, 2003).

35. Stewart, A. J., van Westrenen, W., Schmidt, M. W. & Melekhova, E. Effect of gasketing and assembly design: a novel 10/3.5 mm multi-anvil assembly reaching perovskite pressures. *High Press. Res.* **26**, 293–299 (2006).
36. Campbell, A. J. et al. Pressure–volume–temperature studies of metal–oxide pairs. In *COMPRES Annual Meeting* (2007); ([www.geol.umd.edu/~ajc/Posters/CampbellCOMPRES2007poster.pdf](http://www.geol.umd.edu/~ajc/Posters/CampbellCOMPRES2007poster.pdf)).
37. Predel, B. Fe–Ni in *The Landolt–Börnstein Database* (ed. Madelung, O.) (Springer Materials, 1995); doi:10.1007/10474837\_1321.
38. Wang, J., Lu, X.-G., Sundman, B. & Su, X. Thermodynamic assessment of the Au–Ni system. *Comput. Coupling Phase Diagr. Thermochem.* **29**, 263–268 (2005).
39. Frost, D. J. Fe<sup>2+</sup>–Mg partitioning between garnet, magnesio-wüstite and (Mg,Fe)<sub>2</sub>SiO<sub>4</sub> phases of the transition zone. *Am. Mineral.* **88**, 387–397 (2003).
40. McCammon, C. A., Peyronneau, J. & Poirier, J.-P. Low ferric iron content of (Mg,Fe)O at high pressures and temperatures. *Geophys. Res. Lett.* **25**, 1589–1592 (1998).
41. Frost, D. J., Langenhorst, F. & van Aken, P. Fe–Mg partitioning between ringwoodite and magnesio-wüstite and the effect of pressure, temperature and oxygen fugacity. *Phys. Chem. Miner.* **28**, 455–470 (2001).
42. Frost, D. J. & Langenhorst, F. The effect of Al<sub>2</sub>O<sub>3</sub> on Fe–Mg partitioning between magnesio-wüstite and magnesium silicate perovskite. *Earth Planet. Sci. Lett.* **199**, 227–241 (2002).
43. Seifert, S. & O'Neill, H. St C. Experimental determination of activity–composition relations in Ni<sub>2</sub>SiO<sub>4</sub>–Mg<sub>2</sub>SiO<sub>4</sub> and Co<sub>2</sub>SiO<sub>4</sub>–Mg<sub>2</sub>SiO<sub>4</sub> olivine solid solutions at 1200 K and 0.1 MPa and 1573 K and 0.5 GPa. *Geochim. Cosmochim. Acta* **51**, 97–104 (1987).
44. O'Neill, H., St. C., Canil, D. & Rubie, D. C. Oxide metal equilibria to 2500 °C and 25 GPa: implications for core formation and the light component in the Earth's core. *J. Geophys. Res.* **103**, 12239–12260 (1998).
45. Frost, D. J. The structure and sharpness of (Mg,Fe)<sub>2</sub>SiO<sub>4</sub> phase transformations in the transition zone. *Earth Planet. Sci. Lett.* **216**, 313–328 (2003b).
46. Hirschmann, M. Thermodynamics of multicomponent olivines and the solution properties of (Ni,Mg,Fe)<sub>2</sub>SiO<sub>4</sub> and (Ca,Mg,Fe)<sub>2</sub>SiO<sub>4</sub> olivines. *Am. Mineral.* **76**, 1232–1248 (1991).
47. Hirschmann, M. M. & Ghiorso, M. S. Activities of nickel, cobalt and manganese silicate in magmatic liquids and application to olivine/liquid and to silicate/melt partitioning. *Geochim. Cosmochim. Acta* **58**, 4109–4126 (1994).
48. O'Neill, H. St. C. & Wall, V. J. The olivine–orthopyroxene–spinel oxygen geobarometer, the nickel precipitation curve, and the oxygen fugacity of the Earth's upper mantle. *J. Petrol.* **28**, 1169–1191 (1987).
49. Barin, I., Sauerb, F., Schultze-Rhonhof, E. & Sheng, W. S. *Thermochemical Data of Pure Substances* Parts I and II (Weinheim, 1989).
50. Wood, B. J. & Fraser, D. G. *Elementary Thermodynamics for Geologists* (Oxford Univ. Press, 1976).
51. Mukhopadhyay, B., Basu, S. & Holdaway, M. J. A discussion of Margules-type formulations for multicomponent solutions with a generalized approach. *Geochim. Cosmochim. Acta* **57**, 277–283 (1993).
52. Swartzendruber, L. J. The Fe–Ir (iron–iridium) system. *Bull. Alloy Phase Diagrams* **5**, 48–52 (1984).

# Mapping and analysis of chromatin state dynamics in nine human cell types

Jason Ernst<sup>1,2</sup>, Pouya Kheradpour<sup>1,2</sup>, Tarjei S. Mikkelsen<sup>1</sup>, Noam Shores<sup>1</sup>, Lucas D. Ward<sup>1,2</sup>, Charles B. Epstein<sup>1</sup>, Xiaolan Zhang<sup>1</sup>, Li Wang<sup>1</sup>, Robbyn Issner<sup>1</sup>, Michael Coyne<sup>1</sup>, Manching Ku<sup>1,3,4</sup>, Timothy Durham<sup>1</sup>, Manolis Kellis<sup>1,2\*</sup> & Bradley E. Bernstein<sup>1,3,4\*</sup>

**Chromatin profiling has emerged as a powerful means of genome annotation and detection of regulatory activity. The approach is especially well suited to the characterization of non-coding portions of the genome, which critically contribute to cellular phenotypes yet remain largely uncharted. Here we map nine chromatin marks across nine cell types to systematically characterize regulatory elements, their cell-type specificities and their functional interactions. Focusing on cell-type-specific patterns of promoters and enhancers, we define multicell activity profiles for chromatin state, gene expression, regulatory motif enrichment and regulator expression. We use correlations between these profiles to link enhancers to putative target genes, and predict the cell-type-specific activators and repressors that modulate them. The resulting annotations and regulatory predictions have implications for the interpretation of genome-wide association studies. Top-scoring disease single nucleotide polymorphisms are frequently positioned within enhancer elements specifically active in relevant cell types, and in some cases affect a motif instance for a predicted regulator, thus suggesting a mechanism for the association. Our study presents a general framework for deciphering *cis*-regulatory connections and their roles in disease.**

A major challenge in biology is understanding how a single genome can give rise to an organism comprising hundreds of distinct cell types. Much emphasis has been placed on the application of high-throughput tools to study interacting cellular components<sup>1</sup>. The field of systems biology has exploited dynamic gene expression patterns to reveal functional modules, pathways and networks<sup>2</sup>. Yet *cis*-regulatory elements, which may be equally dynamic, remain largely uncharted across cellular conditions.

Chromatin profiling provides a systematic means of detecting *cis*-regulatory elements, given the central role of chromatin in mediating regulatory signals and controlling DNA access, and the paucity of recognizable sequence signals. Specific histone modifications correlate with regulator binding, transcriptional initiation and elongation, enhancer activity and repression<sup>1,3–6</sup>. Combinations of modifications can provide even more precise insight into chromatin state<sup>7,8</sup>.

Here we apply a high-throughput pipeline to map nine chromatin marks and input controls across nine cell types. We use recurrent combinations of marks to define 15 chromatin states corresponding to repressed, poised and active promoters, strong and weak enhancers, putative insulators, transcribed regions, and large-scale repressed and inactive domains. We use directed experiments to validate biochemical and functional distinctions between states.

The resulting chromatin state maps portray a highly dynamic landscape, with the specific patterns of change across cell types revealing strong correlations between interacting functional elements. We use correlated patterns of activity between chromatin state, gene expression and regulator activity to connect enhancers to likely target genes, to predict cell-type-specific activators and repressors, and to identify individual binding motifs responsible for these interactions.

Our results have implications for the interpretation of genome-wide association studies (GWASs). We find that disease variants frequently coincide with enhancer elements specific to a relevant cell

type. In several cases, we can predict upstream regulators whose regulatory motif instances are affected or target genes whose expression may be altered, thereby suggesting specific mechanistic hypotheses for how disease-associated genotypes lead to the observed disease phenotypes.

## Results

### Systematic mapping of chromatin marks in multiple cell types

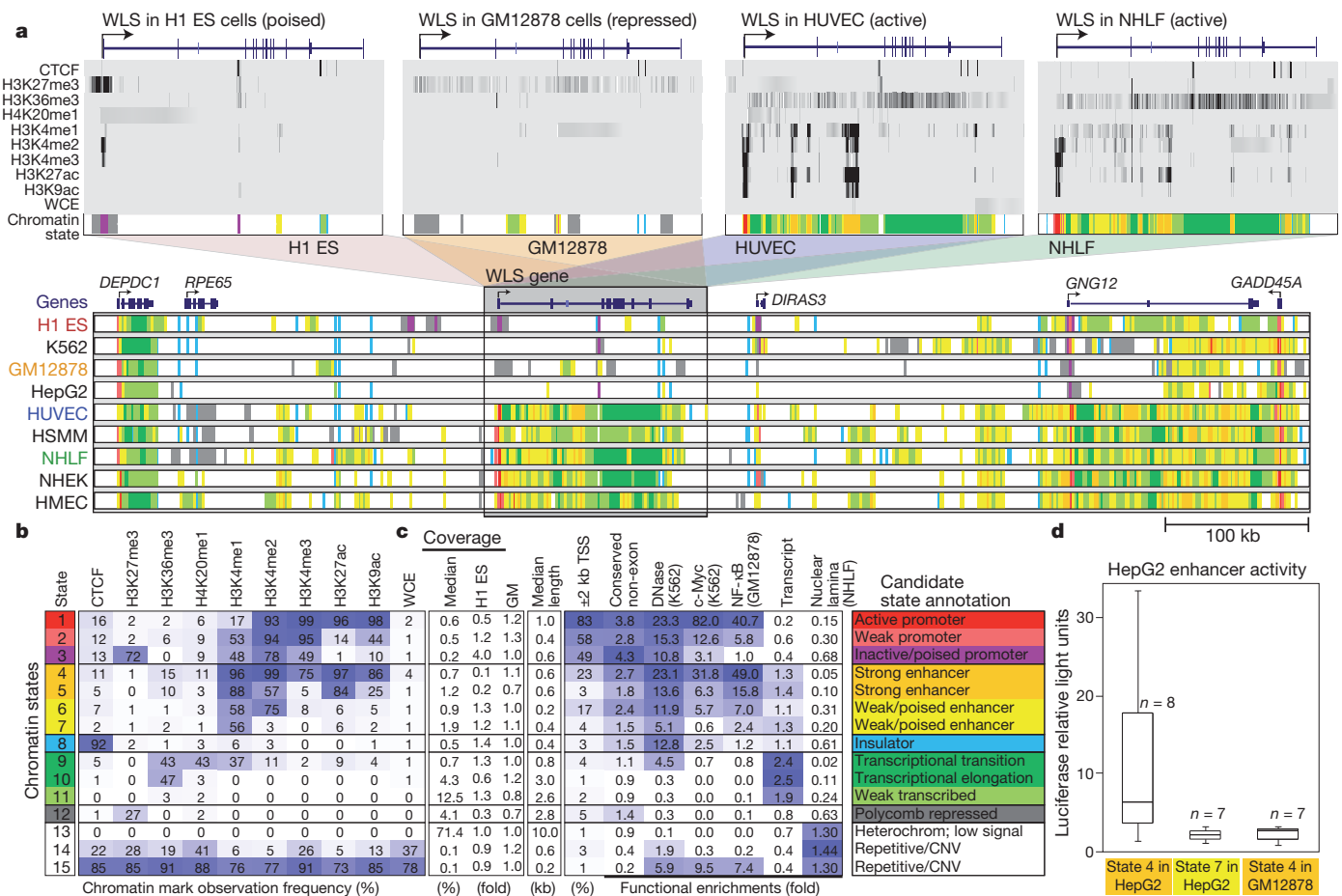
To explore chromatin state in a uniform way across multiple cell types, we applied a production pipeline for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) to generate genome-wide chromatin data sets (Methods and Fig. 1a). We profiled nine human cell types, including common lines designated by the ENCODE consortium<sup>1</sup> and primary cell types. These consist of embryonic stem cells (H1 ES), erythrocytic leukaemia cells (K562), B-lymphoblastoid cells (GM12878), hepatocellular carcinoma cells (HepG2), umbilical vein endothelial cells (HUVEC), skeletal muscle myoblasts (HSMC), normal lung fibroblasts (NHFL), normal epidermal keratinocytes (NHEK) and mammary epithelial cells (HMEC).

We used antibodies for histone H3 lysine 4 trimethylation (H3K4me3), a modification associated with promoters<sup>4,5,9</sup>; H3K4me2 (dimethylation), associated with promoters and enhancers<sup>1,3,6,9</sup>; H3K4me1 (methylation), preferentially associated with enhancers<sup>1,6</sup>; lysine 9 acetylation (H3K9ac) and H3K27ac, associated with active regulatory regions<sup>9,10</sup>; H3K36me3 and H4K20me1, associated with transcribed regions<sup>3–5</sup>; H3K27me3, associated with Polycomb-repressed regions<sup>3,4</sup>; and CTCF, a sequence-specific insulator protein with diverse functions<sup>11</sup>. We validated each antibody by western blots and peptide competitions, and sequenced input controls for each cell type. We also collected data for H3K9me3, RNA polymerase II (RNAPII) and H2A.Z (also known as H2AFZ) in a subset of cells.

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. <sup>2</sup>MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Howard Hughes Medical Institute, Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>4</sup>Center for Systems Biology and Center for Cancer Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

\*These authors contributed equally to this work.





**Figure 1 | Chromatin state discovery and characterization.** **a**, Top: profiles for nine chromatin marks (greyscale) are shown across the WLS gene in four cell types, and summarized in a single chromatin state annotation track for each (coloured according to **b**). WLS is poised in ESCs, repressed in GM12878 and transcribed in HUVEC and NHLF. Its TSS switches accordingly between poised (purple), repressed (grey) and active (red) promoter states; enhancer regions within the gene body become activated (orange, yellow); and its gene body changes from low signal (white) to transcribed (green). These chromatin state changes summarize coordinated changes in many chromatin marks; for example, H3K27me3, H3K4me3 and H3K4me2 jointly mark a poised promoter, whereas loss of H3K27me3 and gain of H3K27ac and H3K9ac mark promoter activation. WCE, whole-cell extract. Bottom: nine chromatin state tracks, one per cell type, in a 900-kb region centred at WLS, summarizing 90 chromatin tracks in directly interpretable dynamic annotations and showing activation and repression patterns for six genes and hundreds of regulatory regions, including enhancer states. **b**, Chromatin states learned jointly across

This resulted in 90 chromatin maps corresponding to ~2,400,000 reads covering ~100,000,000,000 bases across nine cell types, which we set out to interpret computationally.

### Learning a common set of chromatin states across cell types

To summarize these data sets into nine readily interpretable annotations, one per cell type, we applied a multivariate hidden Markov model that uses combinatorial patterns of chromatin marks to distinguish chromatin states<sup>8</sup>. The approach explicitly models mark combinations in a set of ‘emission’ parameters and spatial relationships between neighbouring genomic segments in a set of ‘transition’ parameters (Methods). It has the advantage of capturing regulatory elements with greater reliability, robustness and precision than is possible by studying individual marks<sup>8</sup>.

We learned chromatin states jointly by creating a virtual concatenation of all chromosomes from all cell types. We selected 15 states that showed distinct biological enrichments and were consistently recovered (Fig. 1a, b and Supplementary Fig. 1). Even though states

cell types by a multivariate hidden Markov model. The table shows emission parameters learned *de novo* on the basis of genome-wide recurrent combinations of chromatin marks. Each entry denotes the frequency with which a given mark is found at genomic positions corresponding to the chromatin state. **c**, Genome coverage, functional enrichments and candidate annotations for each chromatin state. Blue shading indicates intensity, scaled by column. CNV, copy number variation; GM, GM12878. **d**, Box plots depicting enhancer activity for predicted regulatory elements. Sequences 250 bp long corresponding either to strong or weak/poised HepG2 enhancer elements or to GM12878-specific strong enhancer elements were inserted upstream of a luciferase gene and transfected into HepG2. Reporter activity was measured in relative light units. Robust activity is seen for strong enhancers in the matched cell type, but not for weak/poised enhancers or for strong enhancers specific to a different cell type. Boxes indicate 25th, 50th and 75th percentiles, and whiskers indicate 5th and 95th percentiles.

were learned *de novo* solely on the basis of the patterns of chromatin marks and their spatial relationships, they showed distinct associations with transcriptional start sites (TSSs), transcripts, evolutionarily conserved non-coding regions, DNase hypersensitive sites<sup>12</sup>, binding sites for the regulators c-Myc<sup>13</sup> (MYC) and NF-κB<sup>14</sup>, and inactive genomic regions associated with the nuclear lamina<sup>15</sup> (Fig. 1c).

We distinguished six broad classes of chromatin states, which we refer to as promoter, enhancer, insulator, transcribed, repressed and inactive states (Fig. 1c). Within them, active, weak and poised<sup>4</sup> promoters (states 1–3) differ in expression level, strong and weak candidate enhancers (states 4–7) differ in expression of proximal genes, and strongly and weakly transcribed regions (states 9–11) also differ in their positional enrichments along transcripts. Similarly, Polycomb-repressed regions (state 12) differ from heterochromatic and repetitive states (states 13–15), which are also enriched for H3K9me3 (Supplementary Figs 2–4).

The states vary widely in their average segment length (~500 base pairs (bp) for promoter and enhancer states versus 10 kb for inactive

regions) and in the portion of the genome covered ( $<1\%$  for promoter and enhancer states versus  $>70\%$  for inactive state 13). For each state, coverage was relatively stable across cell types (Supplementary Fig. 5), with the exception of embryonic stem cells (ESCs) in which the poised promoter state is more abundant but strong enhancer and Polycomb-repressed states are depleted, consistent with the unique biology of pluripotent cells<sup>4,16</sup>.

We confirmed that promoter and enhancer states showed distinct biochemical properties (Supplementary Fig. 6). RNAPII was highly enriched at strong promoters, weakly enriched at strong enhancers and nearly undetectable at weak or poised enhancers, consistent with strong transcription at promoters and reports of weak transcription at active enhancers<sup>17,18</sup>. H2A.Z, a histone variant associated with nucleosome-free regions<sup>19</sup>, was enriched in active promoters and strong enhancers, consistent with nucleosome displacement at TSSs and sites of abundant transcription factor binding in active enhancers.

We also used luciferase reporter assays to validate the functionality of predicted enhancers, the distinction between strong and weak enhancer states, and their predicted cell type specificity. We tested strong enhancers, weak enhancers and strong enhancers specific to an unmatched cell type by transfection in HepG2. We observed strong luciferase activity only for strong enhancer elements from the matched cell type (Fig. 1d).

These results and additional properties of the model (Supplementary Figs 7–10) suggest that chromatin states are an inherent, biologically informative feature of the genome. The framework enables us to reason about coordinated differences in marks by directly studying chromatin state changes between cell types (which we refer to as ‘changes’ or ‘dynamics’ without implying any temporal relationship).

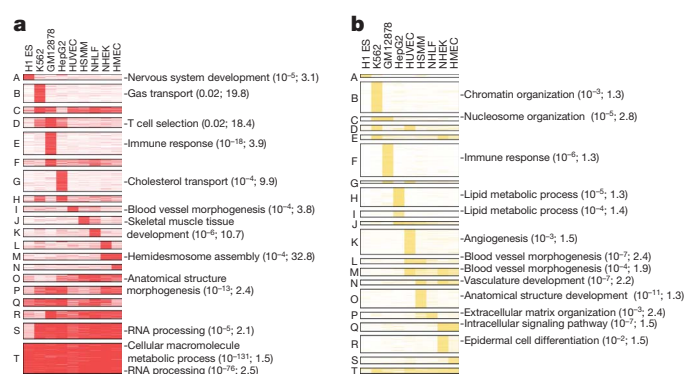
## Extent and significance of chromatin state changes across cell types

We next explored the extent to which chromatin states vary between pairs of cell types. The overall patterns of variability (Supplementary Figs 11 and 12) suggest that regulatory regions vary drastically in activity level across cell types. Enhancer states show frequent interchange between strong and weak, and promoter states vary between active, weak and poised. Promoter states seem more stable than enhancers; they are eight times more likely to remain promoter states, controlling for coverage. Switching was also observed among promoter, enhancer and transcriptional transition states, but no preferential changes to other groups were found. These general patterns suggest that despite varying activity levels, enhancer and promoter regions tend to preserve their chromatin identity as regions of regulatory potential.

Chromatin state differences between cell types relate to cell-type-specific gene functions. An unbiased clustering of chromatin state profiles across annotated TSSs in lymphoblastoid and skeletal muscle cells distinguished informative patterns predictive of downstream gene expression and functional gene classes (Supplementary Figs 13 and 14). Cell-type-specific patterns were also evident when TSSs were simply assigned to the most prevalent chromatin state. Promoters active in skeletal muscle were associated with extracellular structure genes (8.5-fold enrichment), those active in lymphoblastoid cells were associated with immune response genes (7.2-fold enrichment) and those active in both were associated with metabolic housekeeping genes.

### Clustering of promoter and enhancer states on the basis of their activity patterns

Extending our pairwise promoter analysis, we clustered active promoter and strong enhancer regions across all cell types (Methods). This revealed clusters showing common activity and associated with highly coherent functions (Fig. 2). For promoter clusters, these include immune response (GM12878-specific clusters,  $P < 10^{-18}$ ), cholesterol transport (HepG2 specific,  $P < 10^{-4}$ ) and metabolic processes (all cells,  $P < 10^{-13}$ ). Remarkably, genes assigned to enhancer clusters by proximity also showed strong functional enrichments, including immune



**Figure 2 | Cell-type-specific promoter and enhancer states and associated functional enrichments.** **a**, Clustering of genomic locations (rows) assigned to active promoter state 1 (red) across cell types (columns) reveals 20 common patterns of activity (A–T; Methods). For each cluster, enriched gene ontology terms are shown with hypergeometric *P* value and fold enrichment, based on the nearest TSS. For most clusters, several cell types show strong (dark red) or moderate (light red) activity. **b**, Analogous clustering and functional enrichments for strong enhancer state 4 (yellow). Enhancer states show greater cell type specificity, with most clusters active in only one cell type.

response (GM12878 specific,  $P < 10^{-6}$ ), lipid metabolism (HepG2 specific,  $P < 10^{-5}$ ) and angiogenesis (HUVEC specific,  $P < 10^{-3}$ ).

Promoters and enhancers differed in their overall specificities. The majority of promoter clusters showed activity in multiple cell types, consistent with previous work<sup>5,10</sup> (Fig. 2a). Enhancer clusters are significantly more cell type specific, with few regions showing activity in more than two cell types and a majority being specific to a single cell type (Fig. 2b).

We also found differences in the relative contributions of enhancer-based and promoter-based regulation among gene classes. Developmental genes seem to be strongly regulated by both, showing the highest number of proximal enhancers and diverse promoter states, including poised and Polycomb repressed (Supplementary Fig. 15). Tissue-specific genes (for example immune genes and steroid metabolism genes) seem to be more dependent on enhancer regulation, showing multiple tissue-specific enhancers but less diverse promoter states. Lastly, housekeeping genes are primarily promoter regulated, with few enhancers in their vicinities.

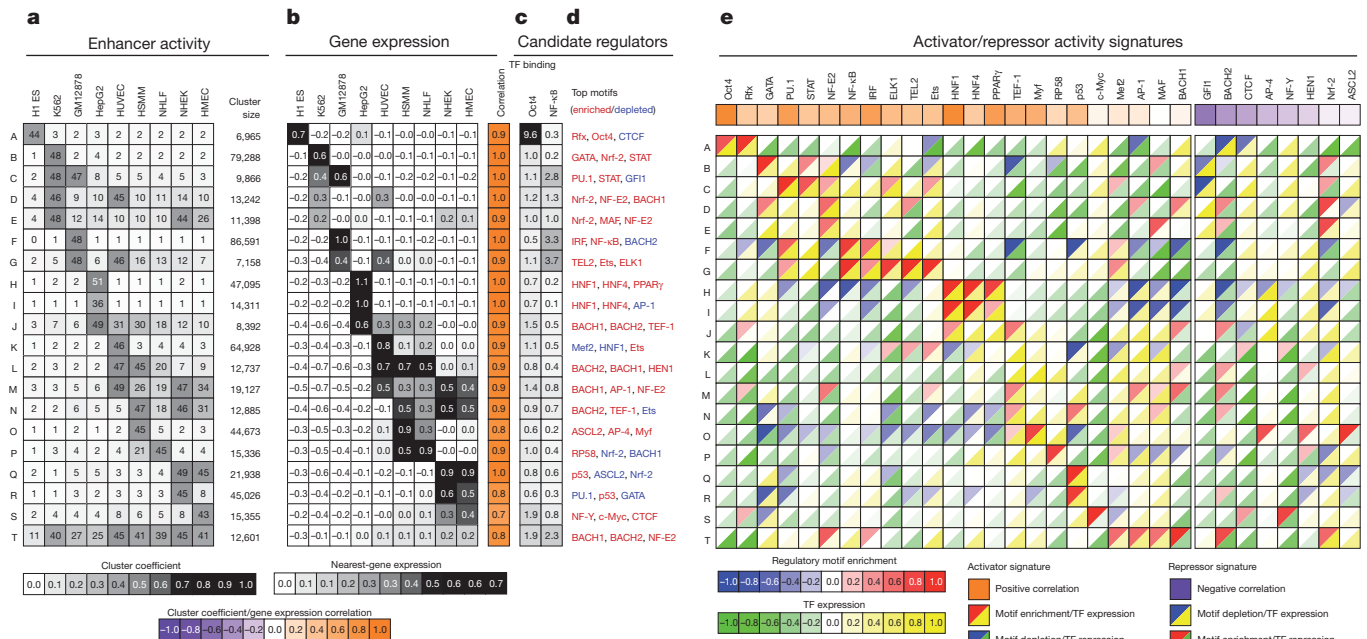
Overall, this dynamic view of the chromatin landscape suggests that multicell chromatin profiles can be as productive for systems biology as expression analysis has traditionally been, and may hold additional information on genome regulatory programs, which we explore next.

## Correlations in activity profiles link enhancers to target genes

We next investigated functional interconnections among enhancers, the factors that activate or repress them, and the genes whose expression they regulate, by defining ‘activity profiles’ for each across the cell types (Fig. 3). We complemented these enhancer activity profiles (Fig. 3a) with profiles for gene expression (Fig. 3b), sequence motif enrichment (Fig. 3d) and the expression of transcription factors recognizing each motif (Fig. 3e). We used correlations between these profiles to probabilistically link enhancers to their downstream targets and upstream regulators (Methods).

We found that patterns of enhancer activity (Figs 2b and 3a) correlated strongly with patterns of nearest-gene expression (Fig. 3b; correlation,  $>0.9$  in 16 of 20 clusters). Because this correlation remained high even for large distances ( $>50$  kb), we used activity correlation as a complement to genomic distance for linking enhancers to target genes (Methods). Activity-based linking yielded an increase in functional gene class enrichment for several clusters (Supplementary Fig. 16).

We validated our approach using quantitative trait locus mapping studies that use covariation between single nucleotide polymorphism (SNP) alleles and gene expression levels to link *cis*-regulatory regions



**Figure 3 | Correlations in activity patterns link enhancers to gene targets and upstream regulators.** **a**, Average enhancer activity across the cell types (columns) for each enhancer cluster (rows) defined in Fig. 2b (labelled A–T) and number of 200-bp windows in each cluster. **b**, Average messenger RNA expression of nearest gene across the cell types and correlation with enhancer activity profile from **a**. High correlations between enhancer activity and gene expression provide a means of linking enhancers to target genes. **c**, Enrichment for Oct4 binding in ESCs<sup>24</sup> and NF-κB binding in lymphoblastoid cells<sup>14</sup> for each cluster. TF, transcription factor. **d**, Strongly enriched (red) or depleted (blue) motifs for each cluster, from a catalogue of 323 consensus motifs. Rfx: Rfx family; Nrf-2: NFE2L2; STAT: STAT family; Ets: Ets family; Mef2: MEF2A and MYEF2;

to target genes. Investigation of four recent quantitative trait locus studies in liver<sup>20</sup> and lymphoblastoid cells<sup>21–23</sup> revealed remarkable agreement with our enhancer predictions. Enhancers linked to a given target gene by our method were significantly enriched for SNPs correlated with the gene's expression level (Supplementary Fig. 17), thus confirming our enhancer–gene linkages with orthogonal data.

### Correlations with transcription factor expression and motif enrichment predict upstream regulators

We next predicted, on the basis of regulatory motif enrichments, sequence-specific transcription factors likely to target enhancers in a given cluster. This implicated a number of transcription factors whose known biological roles matched the respective cell types (Fig. 3d and Supplementary Fig. 18). When ChIP-seq data on the relevant cell type was available, we confirmed that enriched motifs were preferentially bound by the cognate factor (Fig. 3c). Oct4 (POU5F1) motif instances in cluster A (ESC-specific enhancers) were preferentially bound by Oct4 in ESCs<sup>24</sup>, and NF-κB motif instances in cluster F (lymphoblastoid-specific enhancers) were preferentially bound by NF-κB in lymphoblastoid cells<sup>14</sup>. In both cases, motif instances in cell-type-specific enhancers showed a ~5-fold increase in binding in comparison with other enhancers.

However, sequence-based motif enrichments do not distinguish causality. Enrichment could reflect a parallel binding event that does not affect the chromatin state, or the motif could actually be antagonistic to the enhancer state through specific repression in orthogonal cell types. To distinguish between these possibilities, we complemented the enriched motif enrichments with cell-type-specific expression of the corresponding transcription factors (Fig. 3e). We then correlated a 'motif score' based on motif enrichment in a given cluster, and a 'transcription factor expression score' based on the agreement between

Myf: Myf family; NF-Y: NFYA, NFYB and NFYC. **e**, Predicted causal regulators for each cluster based on positive (activators) or negative (repressors) correlations between motif enrichment (top left triangles) and transcription factor expression (bottom right triangles). For example, the red–yellow combination indicates that Oct4 is a positive regulator of ESC-specific enhancers, as its motif-based predicted targets are enriched (red upper triangle) for enhancers active in ESCs (cluster A), and the Oct4 gene is expressed specifically in ESCs, resulting in a positive transcription factor expression correlation (yellow triangle). Overall correlations between motif enrichment and transcription factor expression across all clusters denote predicted activators (positive correlation, orange) and repressors (negative correlation, purple).

the transcription factor expression pattern and the cluster activity profile (Methods). A positive correlation between the two scores implies that the transcription factor may be establishing or reinforcing the chromatin state. A negative correlation would instead imply that the transcription factor may act as a repressor. For example, in addition to the enrichment of the Oct4 motif in the ESC-specific cluster A, Oct4 is specifically expressed in ESCs, leading to the prediction that it is a causal regulator of ESCs (Fig. 3e), consistent with known biology<sup>16</sup>.

For 18 of the 20 clusters, this analysis revealed one or more candidate regulators. Recovery of known roles for well-studied regulators validated our approach. For example, HNF1 (HNF1A), HNF4 (HNF4A) and PPARγ (PPARG) are predicted as activators of HepG2-specific enhancers (clusters H and I), PU.1 (SPI1) and NF-κB as activators of lymphoblastoid (GM12878) enhancers (clusters C, F and G), GATA1 as an activator of K562-specific enhancers (cluster B) and Myf family members as HSMM enhancers<sup>14,25–27</sup> (cluster O).

The analysis also revealed potentially novel regulatory interactions. ETS-related factors (ELK1, TEL2 (ETV7) and Ets family members) are predicted activators of enhancers active in both GM12878 and HUVEC (cluster G) but not of GM12878-specific or HUVEC-specific clusters, emphasizing the value of unbiased clustering. These connections are consistent with reported roles for ETS factors in lymphopoiesis and endothelium<sup>28</sup>. The prediction of p53 (TP53) as an activator in HSMM, NHLF, NHEK and HMEC (clusters N, Q and R) probably reflects its maintained activity in these primary cells, as opposed to cell models in which it may be suppressed by mutation (K562)<sup>29</sup>, viral inactivation (GM12878)<sup>30</sup> or cytoplasmic localization (ESCs)<sup>31</sup>. A widespread role for p53 in regulating distal elements is consistent with its known binding to distal regions<sup>32,33</sup>.

Our analysis also revealed several repressor signatures, including GF11 in K562 and GM12878 (clusters B and C) and BACH2 in ESCs



(cluster A). Both regulators are known to repress transcription by recruiting histone deacetylases and methyltransferases to proximal promoters<sup>34,35</sup>, and GFI1 has also been implicated in the silencing of satellite repeats<sup>35</sup>. Our regulatory inferences suggest that these regulators also modulate chromatin to inhibit enhancer activity, thus suggesting a new mechanism for distal gene regulation.

### Validation of predicted binding events and regulatory outcomes

The regulatory inferences above imply transcription-factor-binding events at motif instances within enhancer regions in specific cellular contexts, and we sought to validate these inferences using a general molecular signature. Binding events are associated with nucleosome displacement, a structural change evident in ChIP-seq data for histones<sup>36</sup>. We thus studied local depletions in the chromatin intensity profiles ('dips') as these are indicative of transcription factor binding. We confirmed that dips were present in individual signal tracks for active enhancers and were associated with preferential sequence conservation and regulatory motif instances (Fig. 4a).

To test our specific predictions, we superimposed chromatin profiles of coordinately regulated enhancer regions, anchoring them on the implied motif instances. Striking dips precisely coincide with regulatory motifs, and are both cell type specific and region specific, exactly as predicted (Fig. 4b, c). Because dips only appear when the factor is expressed, they also support the identity of the *trans*-acting transcription factor.

To confirm that predicted causal motifs contribute to enhancer activity, we used luciferase reporters. Our model implicated HNF regulators as activators of HepG2-specific enhancers (Fig. 3), and context-specific dips supported binding interactions (Fig. 4c). We thus selected for functional analysis ten sites with HNF motifs showing dips in strong HepG2-specific enhancers, and evaluated them with and without the HNF motif. We found that permutation of

the motif consistently led to a reduction in enhancer activity (Fig. 4d), supporting its predicted causal role.

### Assigning candidate regulatory functions to disease-associated variants

Finally, we explored whether our chromatin annotations and regulatory predictions can provide insight into sequence variants associated with disease phenotypes. To that effect, we gathered a large set of non-coding SNPs from GWAS catalogues, an exceedingly small proportion of which are understood at present<sup>37</sup>.

We found that disease-associated SNPs are significantly more likely to coincide with strong enhancers (states 4 and 5; twofold enrichment,  $P < 10^{-10}$ ), despite the fact that no notable association with these states are seen for SNPs in general or for those SNPs tested in the studies. To test whether SNPs associated with a particular disease might have even more specific correspondences, we examined 426 GWAS data sets. We identified ten studies<sup>38–47</sup> whose variants showed significant correspondences to cell-type-specific strong enhancer states (Methods and Fig. 5a).

Individual variants from these studies were strongly enriched in enhancer states specifically active in relevant cell types (Fig. 5a, b). For example, SNPs associated with erythrocyte phenotypes<sup>38</sup> were found in erythrocytic leukaemia cell (K562) enhancers, SNPs associated with systemic lupus erythematosus<sup>39</sup> were found in lymphoblastoid cell (GM12878) enhancers and SNPs associated with triglyceride<sup>40</sup> phenotypes or blood lipid phenotypes<sup>41</sup> were found in hepatocellular carcinoma cell (HepG2) enhancers. We also applied our model to chromatin data for T cells<sup>3</sup> (Supplementary Fig. 19), for which strong enhancer states correlated to variants associated with risk of childhood acute lymphoblastic leukaemia<sup>48</sup>, further validating our approach.

We also used our predicted enhancer/target gene associations to find candidate downstream genes whose expression might be affected by *cis* changes occurring in the enhancer region. Although most of the predicted target genes are proximal to the enhancer, a subset of more distal predicted targets could reflect novel candidates for the disease phenotypes (Fig. 5b).

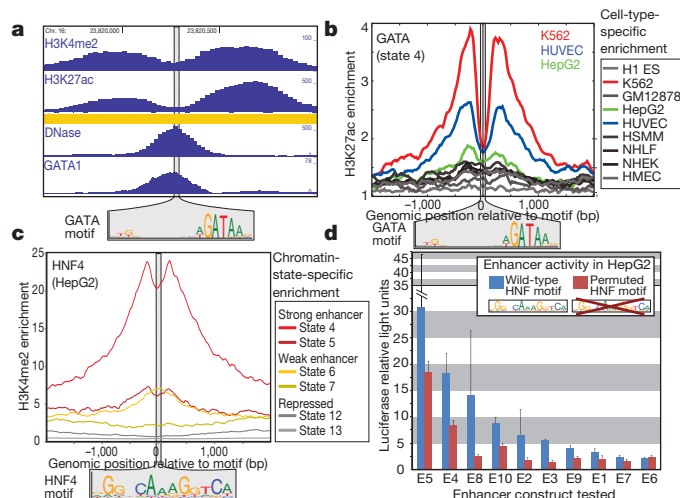
In addition, we identified several instances in which a lead GWAS variant does not correspond to a particular chromatin element but a linked variant coincides with an enhancer with the predicted cell type specificity (Fig. 5c). Thus, chromatin profiles may provide a general means of triaging variants within a haplotype block, a common problem faced in GWASs.

Lastly, we identified several cases in which a disease-associated SNP created or disrupted a regulatory motif instance for a predicted causal transcription factor in the relevant cell type (Fig. 5d), suggesting a specific molecular mechanism by which the disease-associated genotype could lead to the observed disease phenotype consistent with our regulatory predictions.

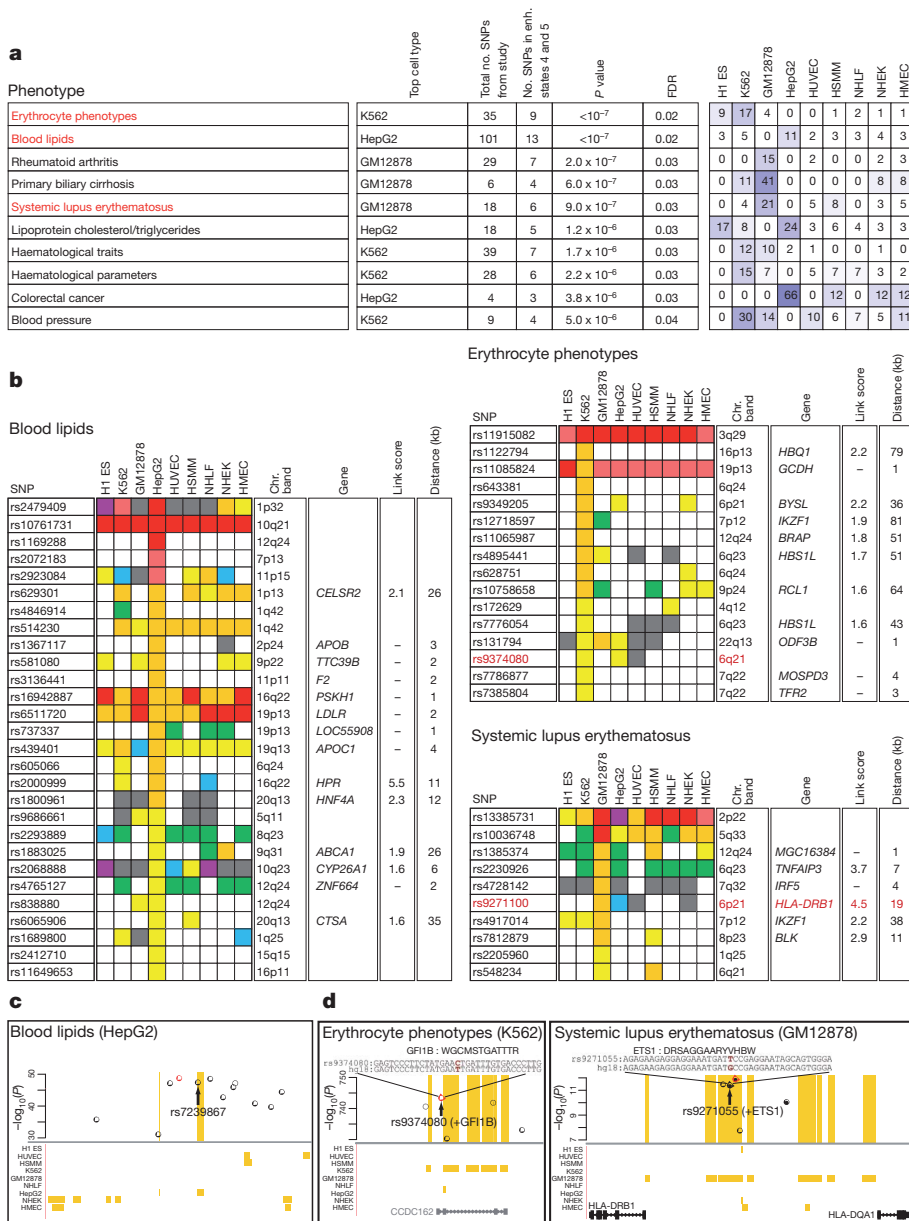
### Discussion

Our work demonstrates the power of multicell chromatin profiles as an additional and dynamic layer of genome annotation. We presented methods to distinguish different classes of functional elements, elucidate their cell type specificities and predict *cis*-regulatory interactions that drive gene expression programs. By intersecting our predictions with non-coding SNPs from GWAS data sets, we proposed potential mechanistic explanations for disease variants, either through their presence within cell-type-specific enhancer states or by their effect on binding motifs for predicted regulators.

Chromatin states drastically reduced the large combinatorial space of 90 chromatin data sets ( $2^{90}$  combinations) to a manageable set of biologically interpretable annotations, thus providing an efficient and robust way to track coordinated changes across cell types. This allowed the systematic identification and comparison of more than 100,000 promoter and enhancer elements. Both types of element are cell type specific, are associated with motif enrichments and assume



**Figure 4 | Validation of regulatory predictions by nucleosome depletions and enhancer activity.** **a**, Dips in chromatin intensity profiles in a K562-specific strong enhancer (orange) coincide with a predicted causal GATA motif instance (logo). The dips probably reflect nucleosome displacement associated with transcription factor binding, supported by DNase hypersensitivity<sup>12</sup> and GATA1 binding<sup>25</sup>. **b**, Superposition of H3K27ac signal across loci containing GATA motifs, centred on motif instances, shows dips in K562, as predicted. **c**, Superposition of H3K4me2 signal for HepG2 shows dips over HNF4 motifs in strong enhancer states, as predicted. **d**, HepG2-specific strong enhancers with predicted causal HNF motifs were tested in reporter assays. Constructs with permuted HNF motifs (red) led to significantly reduced luciferase activity in comparison with wild type (blue), with an average twofold reduction. Data shown are mean luciferase relative light units over three replicates and 95% confidence intervals.



**Figure 5 | Disease variants annotated by chromatin dynamics and regulatory predictions.**  
**a**, Intersection of strong enhancer states (4 and 5) with disease-associated SNPs from GWASs shows significant enrichment (blue shading) in relevant cell types (Methods). Fold enrichments of the SNPs in strong enhancer states for each cell type are indicated. FDR, false-discovery rate. **b**, For three GWAS data sets<sup>38–40</sup>, state annotations are shown for a subset of lead SNPs in the nine cell types (colours as in Fig. 1b, except state 11 is white). The strong enhancer state (orange) is most prevalent in cell types related to the phenotype. For SNPs overlapping strong enhancers, proximal genes with correlated expression are indicated, with linking score and distance. **c**, Example GWAS locus with blood lipid trait<sup>41</sup> association, where the lead variant (red circle) has no functional annotation but a linked SNP (arrow) coincides with a HepG2-specific strong enhancer (orange) and may represent a causal variant. Strong enhancer annotations are shown for all cell types. **d**, Example GWAS loci where a disease SNP affects a conserved instance of a predicted causal motif. Left: lead SNP rs9374080 in the erythrocyte phenotype GWAS<sup>38</sup> is <100 bp from a strong enhancer in K562 and strengthens a motif for GFI1B, a predicted repressor in K562 (Fig. 3e). Right: SNP rs9271055 associated with lupus<sup>39</sup> coincides with a lymphoblastoid (GM12878) strong enhancer and strengthens a motif for ETS1, a predicted activator of lymphoblastoid enhancers (Fig. 3e). This factor is further implicated by lupus-associated variants that directly affect the *ETS1* locus<sup>39</sup>.

strong, weak and poised states that correlate with neighbouring gene expression and function. Enhancers showed very high tissue specificity, enrichment in the vicinity of developmental and cell-type-specific genes, and predictive power for proximal gene expression, reinforcing their roles as sentinels of tissue-specific gene expression<sup>49</sup>. By elucidating enhancers systematically, and linking them to upstream regulators and downstream genes, our analysis can help provide a missing link between regulators and target genes. The power of the approach should increase considerably as additional phenotypically distinct cell types are surveyed, and should enable a greater proportion of enhancer elements to be incorporated into the connectivity network.

The inferred *cis*-regulatory interactions make specific testable predictions, many of which were confirmed through additional experiments and analyses. Our enhancer/target gene linkages are supported by *cis*-regulatory inferences from quantitative trait locus mapping studies. Predicted transcription factor/motif interactions within cell-type-specific enhancers were confirmed in specific cases by transcription factor binding and more generally by depletions in the chromatin profiles at causal motifs in appropriate cellular contexts. Motifs predicted as causal regulators of cell-type-specific enhancers were also confirmed in enhancer assays.

The regulatory inferences afforded by multicell chromatin profiles are unique and highly complementary to data sets for transcription factor binding, expression, chromatin accessibility, nucleosome positioning and chromosome conformation<sup>50</sup>. For example, our regulatory predictions can help focus the spectrum of transcription-factor-binding events to a smaller number of functional interactions. The 'chromatin-centric' approach also complements the extensive body of work on biological network inference from expression data with the potential to introduce enhancers and other genomic elements into connectivity networks.

Our study has important implications for the understanding of disease. Our detailed and dynamic functional annotations of the relatively uncharted non-coding genome can facilitate the interpretation of GWAS data sets by predicting specific cell types and regulators related to specific diseases and phenotypes. Furthermore, the connections derived for enhancer regions, to upstream regulators and downstream genes, suggest *cis*- and *trans*-acting interactions that may be modulated by the sequence variants. Although the present study represents only a first, small step in this direction, we expect that future iterations with a greater diversity of cell types and improved methodologies will help define the molecular underpinnings of human disease.

## METHODS SUMMARY

We performed ChIP-seq analysis in biological replicate as previously described<sup>4</sup>, using antibodies validated by western blots and peptide competitions. ChIP DNA and input controls were sequenced using the Illumina Genome Analyser. Expression profiles were acquired using Affymetrix GeneChip arrays. Chromatin states were learned jointly by applying a hidden Markov model<sup>8</sup> to ten data tracks for each of the nine cell types. We focused on a 15-state model that provides sufficient resolution to resolve biologically meaningful patterns yet is reproducible across cell types when independently processed. We used this model to produce nine genome-wide chromatin state annotations, which were validated by additional ChIP experiments and reporter assays. Multicell type clustering was conducted on locations assigned to strong promoter state 1 (or strong enhancer state 4) in at least one cell type using the *k*-means algorithm. We predicted enhancer/target gene linkages by correlating normalized signal intensities of H3K27ac, H3K4me1 and H3K4me2 with gene expression across cell types as a function of distance to the TSS. Upstream regulators were predicted using a set of known transcription factor motifs assembled from multiple sources. Motif instances were identified by sequence match and evolutionary conservation. We based *P* values for GWAS studies on randomizing the location of SNPs, and based the false-discovery rate on randomizing the assignment of SNPs across studies. Data sets are available from the ENCODE website (<http://genome.ucsc.edu/ENCODE>), the supporting website for this paper ([http://compbio.mit.edu/ENCODE\\_chromatin\\_states](http://compbio.mit.edu/ENCODE_chromatin_states)) and the Gene Expression Omnibus (GSE26386).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 25 August 2010; accepted 4 February 2011.

Published online 23 March 2011.

- Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
- Kim, H. D., Shay, T., O'Shea, E. K. & Regev, A. Transcriptional regulatory circuits: predicting numbers from alphabets. *Science* **325**, 429–432 (2009).
- Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
- Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
- Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
- Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genet.* **39**, 311–318 (2007).
- Hon, G., Wang, W. & Ren, B. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput. Biol.* **5**, e1000566 (2009).
- Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnol.* **28**, 817–825 (2010).
- Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181 (2005).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
- Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
- Raha, D. *et al.* Close association of RNA polymerase II and many transcription factors with Pol III genes. *Proc. Natl Acad. Sci. USA* **107**, 3639–3644 (2010).
- Kasowski, M. *et al.* Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010).
- Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
- Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567–582 (2008).
- De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384 (2010).
- Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010).
- Talbert, P. B. & Henikoff, S. Histone variants — ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11**, 264–275 (2010).
- Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
- Pickrell, J. K. *et al.* Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464**, 768–772 (2010).
- Montgomery, S. B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773–777 (2010).
- Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
- Kunaro, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
- Fujiwara, T. *et al.* Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol. Cell* **36**, 667–681 (2009).
- Lemaigre, F. & Zaret, K. S. Liver development update: new embryo models, cell lineage control, and morphogenesis. *Curr. Opin. Genet. Dev.* **14**, 582–590 (2004).
- Sabourin, L. A. & Rudnicki, M. A. The molecular regulation of myogenesis. *Clin. Genet.* **57**, 16–25 (2000).
- Bartel, F. O., Higuchi, T. & Spyropoulos, D. D. Mouse models in the study of the Ets family of transcription factors. *Oncogene* **19**, 6443–6454 (2000).
- Law, J. C., Ritke, M. K., Yalowich, J. C., Leder, G. H. & Ferrell, R. E. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. *Leuk. Res.* **17**, 1045–1050 (1993).
- Forte, E. & Luftig, M. A. MDM2-dependent inhibition of p53 is required for Epstein-Barr virus B-cell growth transformation and infected-cell survival. *J. Virol.* **83**, 2491–2499 (2009).
- Solozobova, V., Rolletschek, A. & Blattner, C. Nuclear accumulation and activation of p53 in embryonic stem cells after DNA damage. *BMC Cell Biol.* **10**, 46 (2009).
- Cawley, S. *et al.* Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**, 499–509 (2004).
- Wei, C. L. *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* **124**, 207–219 (2006).
- Hoshino, H. *et al.* Co-repressor SMRT and class II histone deacetylases promote Bach2 nuclear retention and formation of nuclear foci that are responsible for local transcriptional repression. *J. Biochem.* **141**, 719–727 (2007).
- Vassen, L., Fiolka, K. & Moroy, T. Gfi1b alters histone methylation at target gene promoters and sites of gamma-satellite containing heterochromatin. *EMBO J.* **25**, 2409–2419 (2006).
- He, H. H. *et al.* Nucleosome dynamics define transcriptional enhancers. *Nature Genet.* **42**, 343–347 (2010).
- Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106**, 9362–9367 (2009).
- Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nature Genet.* **41**, 1191–1198 (2009).
- Han, J. W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genet.* **41**, 1234–1237 (2009).
- Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nature Genet.* **40**, 189–197 (2008).
- Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
- Houlston, R. S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature Genet.* **40**, 1426–1435 (2008).
- Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genet.* **41**, 666–676 (2009).
- Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet.* **42**, 508–514 (2010).
- Liu, X. *et al.* Genome-wide meta-analyses identify three loci associated with primary biliary cirrhosis. *Nature Genet.* **42**, 658–660 (2010).
- Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nature Genet.* **42**, 210–215 (2010).
- Soranzo, N. *et al.* A genome-wide meta-analysis identifies 22 loci associated with eight hematology parameters in the HaemGen consortium. *Nature Genet.* **41**, 1182–1190 (2009).
- Papaemmanuil, E. *et al.* Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. *Nature Genet.* **41**, 1006–1010 (2009).
- Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009).
- Naumova, N. & Dekker, J. Integrating one-dimensional and three-dimensional maps of genomes. *J. Cell Sci.* **123**, 1979–1988 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank members of the epigenomics community at the Broad Institute and the Bernstein and Kellis laboratories, and M. Daly, D. Altschuler and E. Lander for discussions and criticisms. We also thank M. Suva, E. Mendenhall and S. Gillespie for assistance with experiments, and L. Goff and A. Chess for critical reading of the manuscript. We acknowledge the Broad Institute Genome Sequencing Platform for their expertise and assistance with data production. This research was supported by the National Human Genome Research Institute under an ENCODE grant (U54 HG004570; B.E.B.), R01 HG004037 (M. Kellis), RC1 HG005334 (M. Kellis), the Howard Hughes Medical Institute (B.E.B.), the National Science Foundation (awards 0644282 (M. Kellis) and 0905968 (J.E.)) and the Sloan Foundation (M. Kellis).

**Author Contributions** J.E. conducted chromatin state analysis. J.E. and P.K. conducted regulatory motif analysis. J.E. and L.W. conducted GWAS SNP analysis. T.S.M., N.S. and T.D. implemented the ChIP-seq data processing pipeline. C.B.E., X.Z., L.W., R.I., M.C. and M. Ku developed the experimental pipeline and conducted experiments. M. Kellis designed and directed the computational analysis. B.E.B. designed the experimental approach and oversaw the work. J.E., M. Kellis and B.E.B. wrote the paper.

**Author Information** Sequencing and expression data has been deposited into the Gene Expression Omnibus under accession number GSE26386. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to M. Kellis ([manoli@mit.edu](mailto:manoli@mit.edu)).



## METHODS

**Cell culture.** Human H1 ES cells were cultured in TeSR media<sup>51</sup> on Matrigel by Cellular Dynamics International. Cells were split with dispase and collected at a passage number between 30 and 40. Before collection, cells were karyotyped and stained for Oct4 to confirm pluripotency. K562 erythrocytic leukaemia cells (ATCC CCL-243, lot no. 4607240) were grown in suspension in RPMI medium (HyClone SH30022.02) with 10% fetal bovine serum (FBS) and 1% Antibiotic-Antimycotic (GIBCO 15240-062). Cell density was maintained at between  $3 \times 10^5$  and  $7 \times 10^5$  cells ml<sup>-1</sup>. GM12878 B-lymphoblastoid cells (Coriell Cell Repositories, 'expansion A') were grown in suspension in RPMI 1640 medium with 15% FBS (not heat inactivated), 2 mM L-glutamine and 1% penicillin/streptomycin. Cells were seeded at a concentration of  $\sim 2 \times 10^5$  viable cells ml<sup>-1</sup> with minimal disruption, and maintained at between  $3 \times 10^5$  and  $7 \times 10^5$  cells ml<sup>-1</sup>. HepG2 hepatocellular carcinoma cells (ATCC HB-8065, lot no. 4968519) were cultured in DMEM (HyClone SH30022.02) with 10% FBS and 1% penicillin/streptomycin. Cells were trypsinized, resuspended to single-cell suspension, split to a confluence of between 15 and 20% and then collected at  $\sim 75\%$  confluence. NHEK normal human epidermal keratinocytes isolated from skin (Lonza CC-2501, lot no. 4F1155J, passage 1) were grown in keratinocyte basal medium 2 (KGM-2 BulletKit, Lonza) supplemented with BPE, hEGF, hydrocortisone, GA-1000, transferrin, epinephrine and insulin. Cells were seeded at the recommended density (3,500 cells cm<sup>-2</sup>), subjected to two or three passages on polystyrene tissue culture plates and collected at a confluence of 70 to 80%. HSMM primary human skeletal muscle myoblasts (Lonza CC-2580, lot no. 6F4444, passage 2) were cultured in Smooth Muscle Growth Medium 2 (SkGM-2 BulletKit, Lonza) supplemented with rhEGF, dexamethasone, L-glutamine, FBS and GA-1000. Cells were seeded at the recommended density (3,500 cells cm<sup>-2</sup>), subjected to two or three passages and collected at a confluence of 50 to 70%. NHLF primary normal human lung fibroblasts (Lonza CC-2512, lot no. 4F0758, passage 2) were grown in Fibroblast Cell Basal Medium 2 (FGM-2 BulletKit, Lonza) supplemented with hFGF- $\beta$ , insulin, FBS and GA-100. Cells were seeded at the recommended density (2,500 cells cm<sup>-2</sup>), subjected to two or three passages and collected at an approximate confluence of 80%. HUVEC primary human umbilical vein endothelial cells (Lonza CC-2517, lot no. 7F3239, passage 1) were grown in endothelial basal medium 2 (EGM-2 BulletKit, Lonza) supplemented with hFGF- $\beta$ , hydrocortisone, VEGF, R3-IGF-1, ascorbic acid, heparin, FBS, hEGF and GA-1000. Cells were seeded at the recommended density (2,500–5,000 cells cm<sup>-2</sup>), subjected to two or three passages and collected at a confluence of 70 to 80%. HMEC primary human mammary epithelial cells from mammary reduction tissue (Lonza CC-2551, passage 7) were grown in mammary epithelia basal medium (MEGM BulletKit, Lonza) supplemented with hEGF- $\beta$ , hydrocortisone, BPE, GA-1000 and insulin. Cells were seeded at the recommended density (2,500 cells cm<sup>-2</sup>), subjected to two or three passages and collected at 60 to 80% confluence.

**Antibodies.** ChIP assays were performed using the following antibody reagents: H3K4me1 (Abcam ab8895, lot 38311/659352), H3K4me2 (Abcam ab7766, lot 56293), H3K4me3 (Abcam ab8580, lot 331024; Millipore 04-473, lot DAM1623866), H3K9ac (Abcam ab44441, lot 455103/550799), H3K27ac (Abcam ab4729, lot 31456), H3K36me3 (Abcam ab9050, lot 136353), H4K20me1 (Abcam ab9051, lot 104513/519198), H3K27me3 (Millipore 07-449, lot DAM1387952/DAM1514011), CTCF (Millipore 07-729, lot 1350637), H3K9me3 (Abcam ab8898, lot 484088), H2A.Z (Millipore 07-594, lot DAM1504736) and RNAPII N terminus (Santa Cruz sc-899X, lot H0510). All antibody lots were extensively validated for specificity and efficacy in ChIP-seq. Western blots were used to confirm specific recognition of histone protein (or CTCF). Dot plots performed using arrayed histone tail peptides representing various modification states were used to confirm specificity for the appropriate modification. ChIP-seq assays performed on a common cell reagent were used to confirm consistency between different lots of the same antibody.

**Chromatin immunoprecipitation.** Cells were harvested by crosslinking with 1% formaldehyde in cell culture medium for 10 min at 37 °C. After quenching with the addition of 125 mM glycine for 5 min at 37 °C, the cells were washed twice with cold PBS containing protease inhibitor (Roche). After aspiration of all liquid, pellets consisting of  $\sim 10^7$  cells were flash frozen and stored at  $-80$  °C. Fixed cells were thawed and sonicated to obtain chromatin fragments of  $\sim 200$  to 700 bp using a Bioruptor (Diagenode). Immunoprecipitation was performed as previously described, retaining a fraction of input 'whole-cell extract' as a control<sup>4</sup>. Briefly, sonicated chromatin was diluted tenfold and incubated with  $\sim 5$   $\mu$ g antibody overnight. Antibody–chromatin complexes were pulled-down using protein A sepharose, washed and then eluted. After crosslink reversal and proteinase K treatment, immunoprecipitated DNA was extracted with phenol, precipitated in ethanol and treated with RNase. ChIP DNA was quantified by fluorometry using the Qubit assay (Invitrogen).

**Next-generation sequencing.** For each ChIP or control sample,  $\sim 5$  ng of DNA was used to generate a standard Illumina sequencing library. Briefly, DNA fragments were end-repaired using the End-It DNA End-Repair Kit (Epicentre), extended with a 3' 'A' base using Klenow (3'  $\rightarrow$  5' exo-, 0.3 U  $\mu$ l<sup>-1</sup>, NEB), ligated to standard Illumina adapters (75 bp with a 'T' overhang) using DNA ligase (0.05 U  $\mu$ l<sup>-1</sup>, NEB), gel-purified on 2% agarose, retaining products between 275 and 700 bp, and subjected to 18 PCR cycles. These libraries were quantified by fluorometry and evaluated by quantitative PCR or a multiplexed-digital-hybridization-based analysis<sup>52</sup> (NanoString nCounter) to confirm representation and specific enrichment of DNA species. Libraries were sequenced in one or two lanes on the Illumina Genome Analyser using standard procedures for cluster amplification and sequencing by synthesis.

**Expression profiling.** Cytosolic RNA was isolated using RNeasy Columns (Qiagen) from the same cell lots as above. Gene expression profiles were acquired using Affymetrix GeneChip arrays. The data were normalized using the GenePattern expression data analysis package<sup>53</sup>. CEL files were processed by RMA, quantile normalization and background correction. Two replicate expression data sets for each cell type were averaged and log<sub>2</sub>-transformed. Gene-level normalization across cell types was computed by mean normalization.

**Primary processing of sequencing reads.** ChIP-seq reads were aligned to human genome build HG18 with MAQ (<http://maq.sourceforge.net/maq-man.shtml>) using default parameters. All reads were truncated to 36 bases before alignment. Signal density maps for visualization were derived by extending sequencing reads by 200 bp in the 3' direction (the estimated median size of ChIP fragments), and then counting the total number of overlapping reads at 25-bp intervals. Replicate ChIP-seq experiments were verified by comparing enriched intervals as previously described<sup>4</sup>, and were then combined into a single data set. For the hidden Markov model (HMM), density maps were derived by extending sequencing reads by 200 bp in the 3' direction and then assigning them to a single 200-bp window based on the midpoint of the extended read. These maps were then binarized at 200-bp resolution on the basis of a Poisson background model using a threshold of  $10^{-4}$ .

**Joint learning of HMM states across cell types.** To handle data from the nine cell types, we concatenated their genomes to create an extended virtual genome that we used to train the HMM. We applied the model to ten tracks corresponding to the different chromatin marks and input using a multivariate HMM as previously described<sup>8</sup>. Here we used a Euclidean distance for determining initial parameters for the nested initialization step. After the HMM had learned and evaluated a set of roughly nested models, considering up to 25 states, we focused on a 15-state model that provides sufficient resolution to resolve biologically meaningful chromatin patterns and yet is highly reproducible across cell types when independently processed (Supplementary Fig. 7). We used this model to compute the probability that each location is in a given state, and then assigned each 200-bp interval to its most likely state for each cell type. Even though our model focuses on presence/absence frequencies of marks, we found that our states also capture signal intensity differences between high-frequency and low-frequency marks (Supplementary Fig. 9).

**Enrichment analysis.** For each state, enrichments for different annotations were computed at 200-bp resolution with the exception of conservation, which was computed at nucleotide resolution. We used annotations obtained through the UCSC Genome Browser<sup>54</sup> for RefSeq TSSs and transcribed regions<sup>55</sup>, PhastCons<sup>56</sup>, DNase-seq for K562 cells<sup>12</sup>, c-Myc ChIP-seq for K562 cells<sup>13</sup>, NF- $\kappa$ B ChIP-seq for GM12878<sup>14</sup>, Oct4 in ESCs<sup>24</sup> and nuclear lamina<sup>15</sup>. Gene functional group enrichments were determined using STEM<sup>57</sup> and biological process annotations in the Gene Ontology database<sup>58</sup>. *P* values were calculated on the basis of the hypergeometric distribution and corrected for multiple testing using Bonferroni correction.

**Comparing chromatin state assignments between cell types.** For each pair of cell types, the chromatin state assignments at each genomic position were compared. We calculated the frequency with which each pair of states occurred, and normalized this against the expected frequency based on the amount of genome covered by each state. The fold enrichments in Fig. 2a reflect an aggregation across all 72 possible pairs of cell types.

**Pairwise promoter clustering.** Promoters for RefSeq genes were clustered on the basis of the most likely chromatin state assignment across a 2-kb region centred on the TSS. Clustering was performed jointly across GM12878 and HSMM, and was restricted to genes with corresponding Affymetrix expression. Briefly, each promoter was treated as a 330-element binary vector in which each component corresponded to a position along the promoter, cell type and state. Clustering was performed on these vectors using the *k*-means algorithm in MATLAB. Gene expression values were calculated on the basis of the corresponding Affymetrix probe set closest to the TSS.

**Multicell type promoter and enhancer clustering.** Promoter state clustering was performed for all 200-bp intervals assigned to the strong promoter state (state 1)

in at least one cell type. Each interval was represented by a single vector whose components are the estimated probabilities that it be in the strong promoter state for each of the nine cell types, accounting for model assignment uncertainty and biological noise. These were determined from the model posterior probabilities of state assignments and a comparison of state assignments in replicate experimental data. Clustering was performed using the *k*-means algorithm in MATLAB. We found that 20 clusters provided sufficient resolution to distinguish major cell-type-specific patterns. Enhancer state clustering was performed for all 200-bp intervals assigned to strong enhancer state 4 in at least one cell type using identical procedures. For the purposes of display in Fig. 2, the locations were randomly down-sampled. For the purpose of identifying enriched functional gene categories in Fig. 2b, enhancers were linked to the nearest TSS up to 50 kb distant excluding those within 5 kb. Enhancer–gene correspondences based on the nearest gene were used for the expression analysis of distance-based linked genes in Fig. 3b.

**Linking enhancer locations to correlated genes.** To predict linkages between enhancer states and target genes, we combined distance-based information with multicell type correlations between gene expression levels and normalized signal intensities for histone modifications associated with enhancer states (H3K4me1, H3K4me2 and H3K27ac). For each enhancer state (4–7), cell type, and 200-bp interval between 5 kb and 125 kb from the TSS, we trained logistic regression classifiers. The classifiers were trained to use mark intensity/expression correlation values to distinguish real instances of pairs of enhancer states and gene expression values from control pairs based on randomly re-assigning expression values to different genes. So that the classifiers learned a smooth and robust function at each position, we included as part of the training all enhancer state assignments within a 10-kb window centred at the position. The link score for a specific enhancer–gene linkage was defined as the ratio of the corresponding logistic regression classifier probability score to that for the randomized data.

For the evaluation of the expression quantitative trait loci (QTL) analysis, we used a link score threshold of 2.5. The expression QTL data was obtained from the University of Chicago QTL browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>). In the QTL evaluation, each SNP that overlapped a strong enhancer state (4 or 5), was within 125 kb of a TSS, excluding locations within 5 kb, and was associated with a gene for which we had expression data was considered eligible to be supported by our linked predictions. We computed the fraction we observed linked on the basis of our linked predictions relative to the fraction that would be expected to be linked conditioned on knowing the distance distributions of the SNPs relative to the gene TSS.

For the evaluation of linked predictions using the Gene Ontology database, we used the same link score threshold and compared gene assignments against the distance-based assignments defined above. The base set of genes in the enrichment analysis here were all genes that could be linked in at least one cluster.

**Motif and transcription factor analysis.** A database of known transcription factor motifs was collated by combining motifs from TRANSFAC<sup>59</sup> (version 11.3), JASPAR<sup>60</sup> (2010-05-07) and protein-binding microarray data sets<sup>61–63</sup>. Motif instances in non-coding and non-repetitive regions of the genome were identified using these motifs and sequence conservation using a 29-way alignment of eutherian mammal genomes (K. Lindblad-Toh *et al.*, submitted). These were filtered using a significance threshold of  $P < 4^{-8}$  for the motifs<sup>64</sup>, and a confidence level based on conservation. Motifs were linked to corresponding transcription factors using metadata provided by the source. Motif enrichments for chromatin state clusters were computed as ratios to the instances of shuffled motifs, to correct for non-specific conservation and composition. A confidence interval was calculated for each ratio using Wilson score intervals ( $z = 1.5$ ), selecting the most conservative value within the confidence interval. In cases where multiple motif variants were available for the same transcription factor, the one that showed the most variance in enrichment across clusters was selected.

For predicting causal activators and repressors, motif scores and transcription factor expression scores were correlated as follows. Motif scores were calculated as described above. Transcription factor expression scores were calculated for each cluster by correlating the expression of the transcription factor across the cell types with the activity profile of the enhancers in that cluster (defined by the cluster means from the *k*-means clustering). The motif scores and the transcription factor expression scores were then correlated against each other to identify positively and negatively correlated transcription factors.

Transcription factor/motif interactions predicted for strong enhancer states in specific cell types were validated by using the raw ChIP-seq tag enrichments as proxy for nucleosome positioning. For this purpose, sequencing reads were processed as above, except that the middle 75 bp of inferred ChIP fragments were used to derive signal density informative of nucleosome depletion (dips), as previously described<sup>36</sup>. Superposition plots show tag enrichments relative to a uniform background computed on the basis of sequencing depth.

**Quantitative real-time PCR.** Enrichment ratios for RNAPII and H2A.Z ChIPs were determined relative to input chromatin by quantitative real-time PCR using an ABI 7900 detection system, in biological replicate as described previously<sup>65</sup>. Regions used for validation correspond to three different chromatin states, including 13 for state 1 (arbitrarily selected), 11 for state 4 (arbitrarily selected but excluding regions within 2 kb of a state-1 annotation) and 11 for state 7 (arbitrarily selected but excluding regions within 2 kb of a state-1 or state-4 annotation). PCR primers are listed in Supplementary Data 1.

**Functional enhancer assays.** The SV40 promoter was first inserted between the HindIII and NcoI sites of pGL4.10 (Promega). Next, 250-bp sequences from the reference genome (hg18) corresponding to different chromatin states (eight from HepG2 state 4, seven from HepG2 state 7 and seven from GM12878 state 4) were synthesized (GenScript) and then inserted between the two SfiI sites upstream of the SV40 promoter. HepG2 cells were seeded into 96-well plates at a density of  $5 \times 10^4$  cells per well and expanded overnight to ~50% confluency. The cells were then transfected with 400 ng of a pGL4.10-derived plasmid and 100 ng of pGL4.73 (Promega) using Lipofectamine LTX. Firefly and Renilla luciferase activities were measured 24 h post-transfection using Dual-Glow (Promega) and an EnVision 2103 multilabel reader (PerkinElmer), from triplicate experiments. Data are reported as light units relative to a control plasmid. For validation of causal transcription factor motifs, ten sequences of 250 bp corresponding to HepG2-specific strong enhancers (state 4) with dips and HNF motifs were tested as above, and compared with identical sequences except with the HNF motif permuted. Tested enhancer elements are listed in Supplementary Data 1.

**GWAS SNP analysis.** The GWAS variants and SNP coordinates were obtained from the NHGRI catalogue and the UCSC browser<sup>37,54</sup> (October 30, 2010). This set was refined by extending the blood lipid GWAS<sup>41</sup> set to contain all reported SNPs, and by bifurcating the haematological and biochemical traits study<sup>46</sup> into a haematological traits set and a biochemical traits set. We limited our analysis to studies reporting two or more associated SNPs. The variants from each study were intersected with chromatin states from each of the cell types. The reported *P* values were based on the overlap of associated SNPs with strong enhancer states 4 and 5. We controlled for non-independence between proximal SNPs by using a randomization test where SNPs were randomly shifted while preserving relative distance. We then defined an estimated false-discovery rate based on permutations in which SNPs were randomly re-assigned to different studies, and recomputed *P* values. Estimates of false-discovery rates based on these permutations control for multiple testing of studies and cell types and for general non-specific enrichments for states 4 and 5 with GWAS hits. Candidate gene targets were predicted for a subset of variants associated with enhancer states on the basis of the lead cell type using the linking method described above.

**Data access.** Data sets are available from the ENCODE website (<http://genome.ucsc.edu/ENCODE>), the supporting website for this paper ([http://compbio.mit.edu/ENCODE\\_chromatin\\_states](http://compbio.mit.edu/ENCODE_chromatin_states)) and the Gene Expression Omnibus (GSE26386).

51. Ludwig, T. E. *et al.* Feeder-independent culture of human embryonic stem cells. *Nature Methods* **3**, 637–646 (2006).
52. Geiss, G. K. *et al.* Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnol.* **26**, 317–325 (2008).
53. Reich, M. *et al.* GenePattern 2.0. *Nature Genet.* **38**, 500–501 (2006).
54. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
55. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
56. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
57. Ernst, J. & Bar-Joseph, Z. STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* **7**, 191 (2006).
58. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genet.* **25**, 25–29 (2000).
59. Matys, V. *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378 (2003).
60. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W. W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94 (2004).
61. Berger, M. F. *et al.* Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell* **133**, 1266–1276 (2008).
62. Badis, G. *et al.* Diversity and complexity in DNA recognition by transcription factors. *Science* **324**, 1720–1723 (2009).
63. Berger, M. F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nature Biotechnol.* **24**, 1429–1435 (2006).
64. Touzet, H. & Varre, J. S. Efficient and accurate P-value computation for Position Weight Matrices. *Algorithms Mol. Biol.* **2**, 15 (2007).
65. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).

# Low strength of deep San Andreas fault gouge from SAFOD core

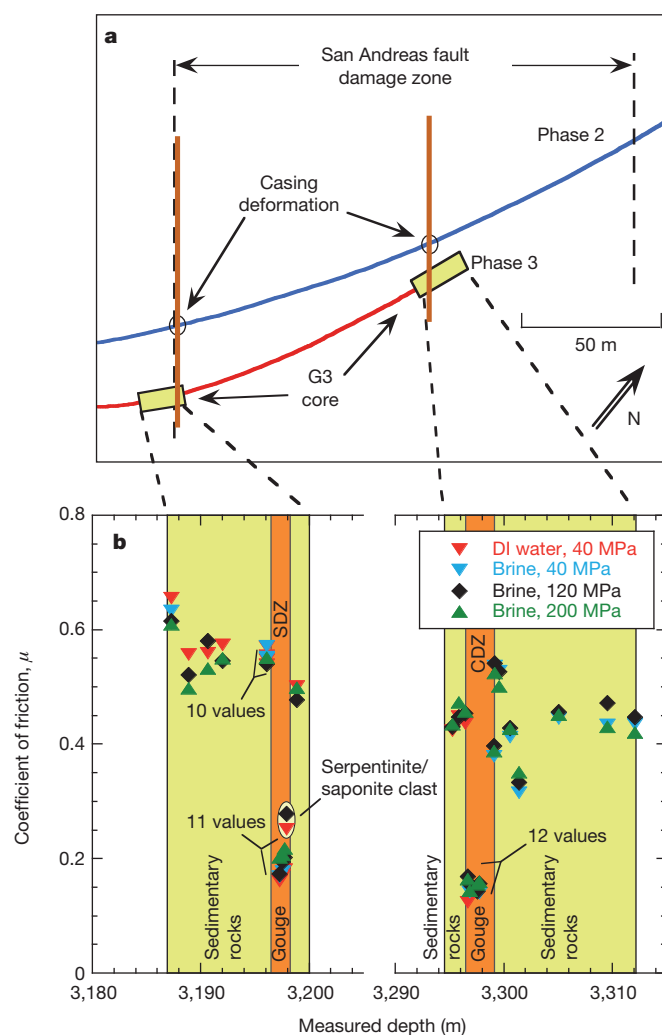
David A. Lockner<sup>1</sup>, Carolyn Morrow<sup>1</sup>, Diane Moore<sup>1</sup> & Stephen Hickman<sup>1</sup>

The San Andreas fault accommodates 28–34 mm yr<sup>-1</sup> of right lateral motion of the Pacific crustal plate northwestward past the North American plate. In California, the fault is composed of two distinct locked segments that have produced great earthquakes in historical times, separated by a 150-km-long creeping zone. The San Andreas Fault Observatory at Depth (SAFOD) is a scientific borehole located northwest of Parkfield, California, near the southern end of the creeping zone. Core was recovered from across the actively deforming San Andreas fault at a vertical depth of 2.7 km (ref. 1). Here we report laboratory strength measurements of these fault core materials at *in situ* conditions, demonstrating that at this locality and this depth the San Andreas fault is profoundly weak (coefficient of friction, 0.15) owing to the presence of the smectite clay mineral saponite, which is one of the weakest phyllosilicates known. This Mg-rich clay is the low-temperature product of metasomatic reactions between the quartzofeldspathic wall rocks and serpentinite blocks in the fault<sup>2,3</sup>. These findings provide strong evidence that deformation of the mechanically unusual creeping portions of the San Andreas fault system is controlled by the presence of weak minerals rather than by high fluid pressure or other proposed mechanisms<sup>1</sup>. The combination of these measurements of fault core strength with borehole observations<sup>1,4,5</sup> yields a self-consistent picture of the stress state of the San Andreas fault at the SAFOD site, in which the fault is intrinsically weak in an otherwise strong crust.

SAFOD is a deep scientific borehole that penetrates the San Andreas fault (SAF) at a vertical depth of approximately 2.7 km and is the deepest land-based scientific drilling project to cross a plate-bounding fault<sup>1,6,7</sup> (see <http://www.earthscope.org> for additional information). During phase 2 drilling in 2005, the basic structure of the SAF was determined (Fig. 1) using borehole logging data<sup>1</sup> and supplementary laboratory studies of the drilling cuttings<sup>8,9</sup>. At 2.7 km depth, the damage zone associated with the fault is approximately 200 m wide, and two actively creeping strands were identified within it by accumulated deformation of the steel casing in the main borehole<sup>1</sup>. These two active shear zones, referred to as the southwest deforming zone (SDZ) and the central deforming zone (CDZ), were primary targets of the phase 3 multilateral core drilling operation in 2007. Approximately 31 m of core were recovered from across the SDZ, CDZ and adjoining damage-zone rocks, including 1.6 m and 2.6 m of highly foliated, incohesive fault gouge associated with the SDZ and CDZ, respectively.

We have completed frictional strength measurements on 25 core samples that span the important lithologic units. Of these, 17 are detrital sedimentary rocks, ranging from fine-grained sandstones to mudstones; representative X-ray diffraction (XRD) patterns are presented in Supplementary Fig. 2. The SDZ and CDZ are represented by four samples apiece. In marked contrast to the adjoining rocks, both foliated gouge zones consist of porphyroclasts of serpentinite and sedimentary rock dispersed in a matrix of Mg-rich clays<sup>10</sup> (Supplementary Fig. 3). XRD patterns of the CDZ were dominated by saponite (estimated to be greater than 60% from petrographic analysis) with some quartz and calcite. The SDZ gouge was composed primarily of saponite + corrensite with some quartz and feldspars (corrensite is a

regularly interlayered chlorite-saponite clay). The porphyroclasts are also partly altered to Mg-rich clays<sup>3</sup>. The two gouge zones are interpreted to be the product of shearing-enhanced metasomatic reactions between serpentinite, tectonically entrained within the fault, and adjoining sedimentary rocks<sup>2,3</sup>.



**Figure 1** | Location and strengths of SAFOD core samples. **a**, Map view of SAF damage zone and SAFOD boreholes from phase 2 (blue) (indicating actively deforming casing) and phase 3 Hole G (red) with location of recovered core (yellow) at approximately 2.7 km vertical depth. Active deformation zones are shown in orange. **b**, Frictional strength of core samples plotted versus measured depth along Hole G (at sliding rate  $V = 1.15 \mu\text{m s}^{-1}$ ). Active fault traces SDZ (3,196.4–3,198.1 m measured depth) and CDZ (3,296.6–3,299.1 m) have notably low strength. A few samples were tested with deionized (DI) water. Extrapolation to SAF plate rate reduces shear zone strength to  $\mu \approx 0.15$ .

<sup>1</sup>US Geological Survey, Menlo Park, California 94025, USA.



Sample strength is reported as coefficient of friction  $\mu = \tau/\bar{\sigma}_n$ , where  $\tau$  and  $\bar{\sigma}_n$  are respectively shear stress and effective normal stress on the test faults; we estimate *in situ*  $\bar{\sigma}_n$  to be  $\sim 122$  MPa (see below). Here,  $\bar{\sigma}_n = \sigma_n - p$ , where  $\sigma_n$  is normal stress and  $p$  is pore pressure. Representative strength tests are plotted in Fig. 2. Frictional strength was compiled from all deformation tests at 9.8–10.4 mm fault-parallel slip and sliding rate  $V = 1.15 \mu\text{m s}^{-1}$ . As shown in Fig. 2, nearly all time- and slip-dependent strengthening had ended by 10 mm slip, so that residual strength is reasonably represented by this value. Residual strength refers to the stable strength of the test sample once fully developed shear flow is established. Because control parameters—including  $\sigma_n$ ,  $p$ ,  $V$  and pore fluid composition—are duplicated in the tests, variations in  $\mu$  are attributed to mineralogical differences between samples. Samples outside the two shear zones show a gradual weakening trend, from  $\mu \approx 0.6$  on the southwestern side to  $\mu \approx 0.4$  on the northeastern side. This trend reflects a compositional change from more quartz-rich sandstone and siltstone on the southwestern side to more phyllosilicate-rich mudstones to the northeast (Supplementary Fig. 2). The SDZ marks the southwestern boundary of the damage zone, so that the samples with the highest residual strength reside outside the damage zone ( $\mu \approx 0.50$ – $0.65$ ).

The most significant strength observation is the abrupt decrease in  $\mu$  within the two actively deforming shear zones. All residual strength measurements of the foliated gouge (at 10.4 mm slip) yield  $\mu \leq 0.21$ ; the weakest sample has a strength of  $\mu = 0.13$  (Fig. 1). A partly altered serpentinite porphyroclast from the SDZ has a strength of  $\mu \approx 0.26$  and apparently survived by weaker matrix material flowing around it. The very low measured strengths are attributed to the abundance of the extremely weak mineral saponite ( $\mu \approx 0.05$ ) (Fig. 2). Petrographic analysis indicates a saponite volume fraction of 60–65% in the foliated gouge matrix. Corrensite was also found in the SDZ and, based on

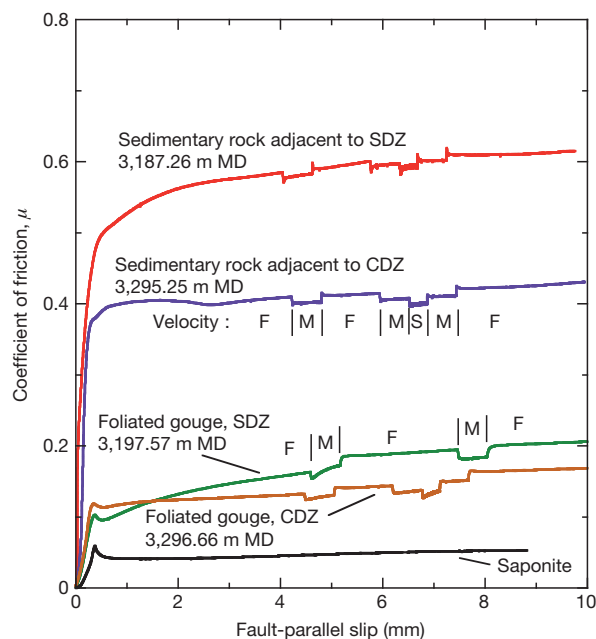
composition, is likely to have a strength of  $0.05 < \mu < 0.4$ . Thus, corrensite along with increased quartz content (Supplementary Fig. 2) may be responsible for the marginally stronger frictional strength of the SDZ. Serpentine and minor amounts of other phyllosilicates (including chlorite, illite and micas) are present in the foliated gouge and, when added to the stronger quartz, feldspar and calcite constituents, result in a matrix strength that is consistent with estimates suggested by mixing law studies<sup>11–13</sup>. Rock fabric that localizes weak minerals can lower frictional strength relative to the strength of ground and mixed samples<sup>14</sup>. Thus, the SAFOD foliated gouge may be even weaker in its undisturbed state than the values reported here.

Shear strength of fault gouge material typically varies with sliding rate. Rate dependence can be important in determining deformation mode (stable or unstable) and in extrapolating shear strength to SAF deformation rates. Steady-state rate sensitivity is defined<sup>15</sup> by the parameter  $(a - b) = d\mu_{ss}/d\ln V$ , where  $\mu_{ss}$  is the steady-state friction coefficient at velocity  $V$ . Imposed velocity steps, as shown in Fig. 2, are used to determine  $a - b$ . Negative values promote unstable slip, whereas positive values are likely to result in stable creep. The serpentinite porphyroclast from the SDZ shows a range of both negative and positive values ( $a - b = +0.0004 \pm 0.0014$ ) similar to past values reported for serpentinites<sup>16,17</sup>. All other core samples have positive rate sensitivity. Samples taken from outside the foliated gouge zones have values in the range  $+0.001 < a - b < +0.007$ . For CDZ, the combined measurements resulted in  $a - b = +0.0018 \pm 0.0008$ . For SDZ, values are twice as large:  $a - b = +0.0037 \pm 0.0007$ .

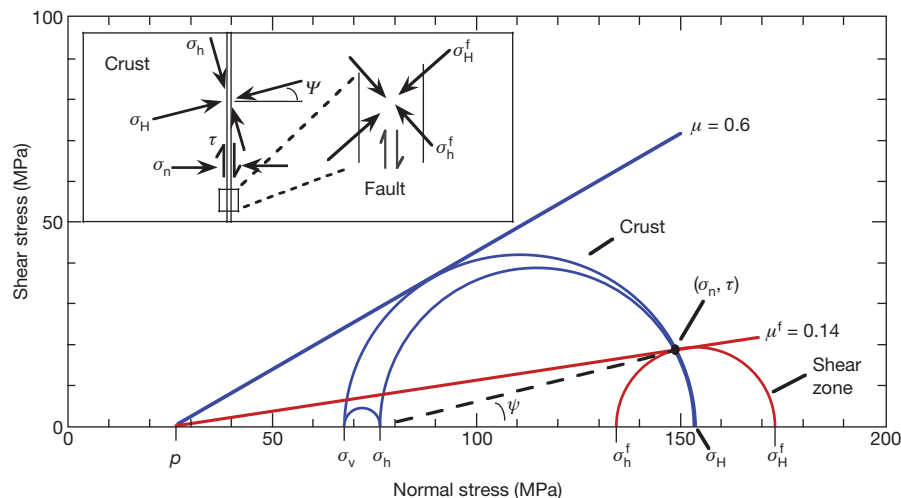
Average *in situ* strengths for CDZ and SDZ gouges (Fig. 1) are  $\mu = 0.16$  and  $0.19$ , respectively. These measurements are determined for a slip rate of  $1.15 \mu\text{m s}^{-1}$  ( $36,000 \text{ mm yr}^{-1}$ ) and should be reduced to the appropriate *in situ* deformation rate ( $\leq 34 \text{ mm yr}^{-1}$ ) of the SAF. (Note that the slowest imposed deformation rate in the strength tests ( $0.0115 \mu\text{m s}^{-1}$ , Fig. 2) is only about 11 times the *in situ* rate.) Extrapolation of test strengths using the observed rate sensitivity for the foliated gouge indicates upper bounds for steady-state strength of CDZ and SDZ, respectively, of  $\mu = 0.14$  and  $0.16$ . This scaling assumes that the slip rate across the 1-mm-thick test gouge layer should be compared to the SAF deformation rate that is accommodated by the combined thickness of the SDZ and CDZ ( $\sim 4.2 \text{ m}$ ). Depending on how strain is partitioned within the shear zones, the actual *in situ* shear strength supported by the deforming zones could be much less.

Although the SAF is one of the most well-studied fault systems in the world, fundamental questions about its strength and mechanical properties remain unanswered<sup>18</sup>. The SAF heat flow paradox was identified more than 40 years ago and is debated to this day<sup>19–23</sup>. Essentially, if the shear strength of the SAF were consistent with common laboratory-derived Byerlee rock friction ( $\mu > 0.6$ ), frictional heating of the fault during earthquakes and stable fault creep should result in increased temperature and heat flow adjacent to the fault zone. In addition, the maximum horizontal stress near the fault should be oriented at  $\sim 30^\circ$  to the fault trace. However, no evidence of a heat-flow anomaly along the creeping section of the SAF has been found<sup>20,24</sup>, and borehole stress observations at SAFOD confirm that the maximum horizontal stress at this locality is at a high angle to the fault trace<sup>4,5,25</sup>. Although formation fluid pressure is apparently above hydrostatic in the sedimentary sequence northeast of the fault, there is no evidence from SAFOD drilling operations that pore pressure within the fault zone is elevated relative to the country rock<sup>1</sup>. The direct measurement, reported here, of low frictional strength ( $\mu \approx 0.15$ ) of foliated gouge material taken at depth from the actively deforming shear zones is consistent with both the lack of an observed heat flow anomaly and the maximum compressive stress oriented at a high angle to the fault trace. Also, the positive dependence of strength on slip rate of the fault gouge material is consistent with deformation by creep rather than by earthquakes.

Saponite becomes unstable above about  $150^\circ\text{C}$  (ref. 26) and is unlikely to be found deeper in the fault zone than 3.5–4 km (observed



**Figure 2 | Four representative deformation tests of core material from Hole G, with saponite for comparison.** Periodic strength steps are due to decade changes in sliding rate (fault velocity in  $\mu\text{m s}^{-1}$ : fast (F), 1.15; medium (M), 0.115; slow (S), 0.0115). Strength variations are attributed to compositional differences between samples as shown in XRD patterns. Bottom curve shows strength of monomineralic saponite gouge taken from vesicles in altered volcanic rocks from the Isle of Skye, Scotland (obtained from Mineralogical Research Co.). Permanent strengthening during some slow velocity steps is the result of time-dependent compaction. Foliated gouge is 3–4 times stronger than pure saponite owing to the presence of strong minerals like quartz. MD, measured depth.



**Figure 3 | Stress state for SAFOD drill site at 2.7 km depth.** Model is based on borehole observations (assuming hydrostatic pore pressure) and foliated gouge strength (blue, host rock; red, weak shear zone). Main panel shows relationship of stress states within and outside the weak shear zone. Inset shows the corresponding spatial orientation of horizontal stresses in the model. While

temperature within the fault at  $\sim 2.7$  km depth at SAFOD was 110–115 °C; C. Williams, personal communication). Stable creep and low strength of the deep SAF in the creeping section may reflect the presence of other low-strength minerals, elevated fluid pressure, or enhanced chemical weakening at greater depth than penetrated by SAFOD. Still, when considering the mechanics of the SAF specifically at 2.7 km, at SAFOD, mineralogy alone appears sufficient to explain fault strength.

Rice<sup>27</sup> analysed the stress state of a weak fault (due to elevated pore pressure,  $p$ ) embedded in a stronger crust in a transpressional regime. Tembe *et al.*<sup>28</sup> extended the Rice analysis to include fault gouge of arbitrary strength. We use  $\sigma_H$  and  $\sigma_h$  for maximum and minimum horizontal principal stresses, respectively, and denote values within the fault by a superscript 'f'. Following Tembe *et al.*, a stress diagram representative of the SAFOD site at 2.7 km depth, with  $\mu^f = 0.14$ ,  $\mu = 0.60$  and hydrostatic  $p$ , is shown in Fig. 3. The model requires that  $\sigma_H$  (outside the weak shear zones) makes an angle of 77° to the strike of the SAF and is consistent with borehole observations at SAFOD showing high differential stresses in a transitional strike-slip to reverse-faulting stress regime, with  $\sigma_H$  maintaining a high angle to the SAF (70–80°) at depth<sup>4,5</sup>. Traction on the fault, in this model, are  $\tau = 17$  MPa and  $\bar{\sigma}_n = 122$  MPa.

SAFOD phase 3 drilling has provided, for the first time, continuous core samples from the actively deforming SAF at a depth of 2.7 km. A self-consistent picture is emerging about the strength and deformation processes of this complex portion of the SAF that represents the transition from locked to creeping portions of the fault. Measurements of frictional strength of core material from within the SAF damage zone show two low-strength ( $\mu \approx 0.15$ ) foliated gouge zones that are 1.6 and 2.6 m wide. These zones correspond to the actively creeping shear zones that were independently identified by casing deformation within the phase 2 hole. These shear zones are embedded in stronger material with  $\mu \approx 0.35$ –0.65. The extremely low strength of the foliated gouge in an otherwise strong crust is sufficient to explain the observed orientation of maximum compressive stress at a high angle relative to the strike of the fault (Fig. 3) without invoking high fluid pressure or other proposed fault-weakening mechanisms.

## METHODS SUMMARY

We measured frictional strength of 25 samples obtained from the SAFOD phase 3 Hole G core, composed of material that could be carved from the core or removed as chips or rubble. Some portions of the rock mass bounding the shear zones had sufficient cohesion to be sampled as solid mini-cores or prisms and will be reported on later. Samples were ground to a powder ( $<150 \mu\text{m}$  diameter) and

maximum shear stress in host rock is high, principal stresses rotate within the shear zone to accommodate the weaker material. In the fault, mean stress is high but shear stress is low. In model:  $\sigma_H = 153$  MPa;  $\sigma_h = 76$ ;  $\sigma_v = 67.5$ ;  $p = 27$ . See text for definitions of symbols used.

sheared in 1- or 2-mm-thick gouge layers between 25.4-mm-diameter driving blocks in a triaxial deformation apparatus<sup>13</sup> (Supplementary Fig. 1). Most samples were saturated with brine equivalent to *in situ* pore fluid (Y. Kharaka and J. Thordsen, personal communication) at 1 MPa constant pore pressure; a few tests were conducted with deionized water. Tests were performed at room temperature, constant effective normal stress (40, 120 and 200 MPa) and constant sliding rate (0.0115, 0.115 and 1.15  $\mu\text{m s}^{-1}$ ). Tests were carried out to 200 MPa to be applicable to *in situ* stress conditions (Fig. 3) and to allow for interpolation to other depths.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 19 October 2010; accepted 31 January 2011.

Published online 23 March 2011.

- Zoback, M., Hickman, S. & Ellsworth, W. Scientific drilling into the San Andreas Fault zone. *Eos* **91**, 197–199 (2010).
- Moore, D. E. & Rymer, M. J. Talc-bearing serpentinite, and the creeping section of the San Andreas fault. *Nature* **448**, 795–797 (2007).
- Moore, D. E. & Rymer, M. J. Metasomatic origin of fault gouge comprising the two creeping strands at SAFOD. *Eos* (Fall suppl.), paper T41A-2105 (2010).
- Boness, N. & Zoback, M. D. A multi-scale study of the mechanisms controlling shear velocity anisotropy in the San Andreas Fault Observatory at Depth. *Geophysics* **71**, F131–F146 (2006).
- Hickman, S. & Zoback, M. D. Stress measurements in the SAFOD pilot hole: implications for the frictional strength of the San Andreas fault. *Geophys. Res. Lett.* **31**, L15S12, doi:10.1029/2004GL020043 (2004).
- Zoback, M. D., Hickman, S. & Ellsworth, W. in *Treatise on Geophysics* Vol. 4 (ed. Schubert, G.) 649–674 (Elsevier, 2007).
- Tobin, H., Ito, H., Behrmann, J., Hickman, S. H. & Kimura, G. in *Report from IODP/ICDP Workshop on Fault Zone Drilling, Scientific Drilling* (eds Ito, H. *et al.*) 5–16 (Special Issue No. 1, Integrated Ocean Drilling Program, Hokkaido, 2007).
- Solum, J. G. *et al.* Mineralogical characterization of protolith and fault rocks from the SAFOD main hole. *Geophys. Res. Lett.* **33**, L21314, doi:10.1029/2006GL027285 (2006).
- Tembe, S. *et al.* Frictional strength of cuttings and core from SAFOD drillhole phases 1 and 2. *Geophys. Res. Lett.* **33**, L23307, doi:10.1029/2006GL027626 (2006).
- Holdsworth, R. E. *et al.* Fault rocks from the SAFOD core samples: implications for weakening at shallow depths along the San Andreas Fault, California. *J. Struct. Geol.* **33**, 132–134 (2011).
- Moore, D. E. & Lockner, D. A. Frictional strengths of talc-serpentinite and talc-quartz mixtures. *J. Geophys. Res.* **116**, B01403, doi:10.1029/2010JB007881 (2011).
- Crawford, B. R., Faulkner, D. R. & Rutter, E. H. Strength, porosity, and permeability development during hydrostatic and shear loading of synthetic quartz-clay fault gouge. *J. Geophys. Res.* **113**, B03207, doi:10.1029/2006JB004634 (2008).
- Tembe, S., Lockner, D. A. & Wong, T.-f. Effect of clay content and mineralogy on frictional sliding behavior of simulated gouges: binary and ternary mixtures of quartz, illite and montmorillonite. *J. Geophys. Res.* **115**, B03416, doi:10.1029/2009JB006383 (2010).
- Collettini, C., Niemeijer, A., Viti, C. & Marone, C. Fault zone fabric and fault weakness. *Nature* **462**, 907–910 (2009).
- Dieterich, J. H. Modeling of rock friction 1. Experimental results and constitutive equations. *J. Geophys. Res.* **84**, 2161–2168 (1979).

16. Moore, D. E., Lockner, D. A., Ma, S., Summers, R. & Byerlee, J. D. Strengths of serpentinite gouges at elevated temperatures. *J. Geophys. Res.* **102**, 14787–14801 (1997).
17. Moore, D. E., Lockner, D. A., Tanaka, H. & Iwata, K. The coefficient of friction of chrysotile gouge at seismogenic depths. *Int. Geol. Rev.* **46**, 385–398 (2004).
18. Hickman, S. H. Stress in the lithosphere and the strength of active faults. *Rev. Geophys.* **29**, 759–775 (1991).
19. Brune, J. N., Henyey, T. L. & Roy, R. F. Heat flow, stress, and rate of slip along the San Andreas fault, California. *J. Geophys. Res.* **74**, 3821–3827 (1969).
20. Lachenbruch, A. H. & Sass, J. H. Heat flow and energetics of the San Andreas fault zone. *J. Geophys. Res.* **85**, 6185–6222 (1980).
21. Scholz, C. H. Evidence for a strong San Andreas fault. *Geology* **28**, 163–166 (2000).
22. Zoback, M. D. Strength of the San Andreas. *Nature* **405**, 31–32 (2000).
23. Saffer, D. M., Bekins, B. A. & Hickman, S. Topographically driven groundwater flow and the San Andreas heat flow paradox revisited. *J. Geophys. Res.* **108**, 2274, doi:10.1029/2002JB001849 (2003).
24. Lachenbruch, A. H. & Sass, J. H. Heat flow from Cajon Pass, fault strength, and tectonic implications. *J. Geophys. Res.* **97**, 4995–5015 (1992).
25. Zoback, M. D. *et al.* New evidence on the state of stress of the San Andreas fault system. *Science* **238**, 1105–1111 (1987).
26. Inoue, A. & Utada, M. Smectite-to-chlorite transformation in thermally metamorphosed volcanoclastic rocks in the Kamikita area, northern Honshu, Japan. *Am. Mineral.* **76**, 628–640 (1991).
27. Rice, J. R. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-f.) 475–503 (Academic, 1992).
28. Tembe, S., Lockner, D. & Wong, T.-f. Constraints on the stress state of the San Andreas fault with analysis based on core and cuttings from San Andreas Observatory at Depth (SAFOD) drilling phases 1 and 2. *J. Geophys. Res.* **114**, B11401, doi:10.1029/2008JB005883 (2009).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Weymer, J. Firth and P. Nelson for their assistance and patience in providing the core material. Comments from W. Ellsworth, D. Faulkner and R. Simpson significantly improved the manuscript.

**Author Contributions** C.M. and D.A.L. conducted deformation experiments, D.M. provided petrographic analysis and XRD measurements, and S.H. provided borehole logging results and related stress analysis.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to D.A.L. ([dlockner@usgs.gov](mailto:dlockner@usgs.gov)).



## METHODS

**Sample preparation.** We measured frictional strength of 25 samples that were obtained from the SAFOD phase 3 Hole G core. Hole G was cored in measured depth intervals 3,186.7–3,199.5 m and 3,294.9–3,312.7 m to sample localized shear zones within the SAF damage zone that correspond to the two intervals, referred to as SDZ and CDZ, where slow deformation was observed in the Phase 2 casing<sup>1</sup>. As indicated in Fig. 1, the cored intervals in Hole G were within or adjacent to the SAF damage zone as determined by logging data following phase 2 drilling. While selected samples were obtained to provide whole, undisturbed wafers for intact strength tests, samples tested here were either carved from the core or collected as rubble, chips or loose powder. All samples reported here were prepared by repeated gentle grinding with mortar and pestle until all material passed through a 100 mesh sieve (0.15 mm diameter). Resulting powder was then wetted to make a paste that was formed into a 1-mm-thick test layer (2-mm layers were used in the 200 MPa tests). The first 14 tests were performed with deionized water. All remaining tests used a prepared brine solution that duplicates the major cations and their relative concentrations found in the formation fluid retrieved from the SAFOD drill hole on the northeastern side of the SAF (Y. Kharaka and J. Thordsen, personal communication). Test fluid constituents, expressed in units of grams per litre, are:  $\text{Cl}^-$ , 13.32;  $\text{Na}^+$ , 5.34;  $\text{Ca}^{2+}$ , 2.77; and  $\text{K}^+$ , 0.22. Comparison tests showed only a slight difference between frictional strength for samples sheared with deionized water and samples sheared with the brine solution. Before mechanical testing, X-ray diffraction patterns were obtained to determine mineral composition and relative abundance.

**Testing details.** Tests were performed in a standard triaxial apparatus at room temperature and effective normal stresses of 40, 120 and 200 MPa. A constant pore pressure of 1 MPa was applied in all tests. Samples were 25.4-mm-diameter right-cylinders that contained a sawcut inclined  $30^\circ$  to the sample axis (Supplementary Fig. 1). The sawcut forcing blocks were sandstone–sandstone, sandstone–granite, or granite–granite pairs. Surfaces of forcing blocks were roughened with 100 grit abrasive, to assure good frictional contact with the applied gouge layer. See, for example, refs 9 and 13 for details. Berea sandstone forcing blocks have high permeability but also have 20% porosity that decreases with applied load. The standard test geometry used Berea for the top forcing block to assure rapid hydraulic communication of the fault with the external pore fluid system. Pore fluid flow in and out of the lower driving block was through the fault gouge. To minimize pore pressure transients due to stress changes, low-porosity granite was

used for the lower driving block in most tests, and particularly in experiments with low-permeability clay-rich gouge.

A greased Teflon shim, placed between the piston and the sample assembly, allowed easy lateral slip of the lower driving block during shearing (Supplementary Fig. 1). Samples were jacketed in 3.2-mm-thick latex tubing for isolation from confining fluid. Separate calibration tests showed that the latex tubing provided an equivalent shear resistance of  $0.043 \text{ MPa mm}^{-1}$  due to stretching during deformation experiments. This shear strength correction has been applied to all test results. Shear and normal stresses have also been corrected for the reduction in contact area as the two sample halves slide past each other<sup>13</sup>. Axial load was measured with an internal load cell. Axial shortening, confining pressure and pore pressure were all measured at 1-s intervals. Shear and normal stress resolved on the fault surface were computed in real time from the axial stress, confining pressure and axial shortening. As sample strength varied, confining pressure was adjusted every second to maintain constant normal stress. Axial stress, confining pressure and pore pressure have accuracies of at least 0.03 MPa. Samples were sheared to 9 mm axial shortening ( $\sim 10.4 \text{ mm}$  parallel to the sawcut) at axial shortening rates of 0.01, 0.1 and  $1.0 \mu\text{m s}^{-1}$  to determine the dependence of shear strength on sliding rate and thereby the tendency for stable creep or unstable slip. Slip and slip rate on the inclined fault surfaces were 15% higher than the corresponding axial values. Steady-state changes in strength were estimated for individual velocity steps by measuring the residual strength change after de-trending the friction–displacement curves for long-term strain hardening. This procedure was carried out manually.

**Measurement accuracy.** Sample strength is reported as coefficient of friction  $\mu = \tau/\sigma_n$ . Within a single experiment, changes in  $\mu$  have a precision of  $\pm 0.001$ . Reproducibility of  $\mu$  between experiments, including variations due to sample preparation, is approximately  $\pm 0.005$ . Accuracy of  $\mu$ , after corrections for true contact area and jacket strength, is approximately  $\pm 0.01$ . Initial gouge layer thickness is 1.0 mm for 40 and 120 MPa tests. Gouge layer thickness is 2.0 mm for 200 MPa tests to offset thinning at the higher normal stress. Compaction is not measured during experiments, but layer thickness following experiments is reduced by 5–30%, depending on normal stress, gouge clay content and driving block type. As shear will localize within the gouge layer to different degrees and at different times, depending on composition and normal stress, estimates of true shear strain are problematic. The deformation quantity that is most accurately determined in these experiments is total fault-parallel slip. This can be converted to a nominal shear strain by dividing by the initial 1 or 2 mm gouge thickness.

# *Shank3* mutant mice display autistic-like behaviours and striatal dysfunction

João Peçã<sup>1,2\*</sup>, Cátia Feliciano<sup>1,3\*</sup>, Jonathan T. Ting<sup>1</sup>, Wenting Wang<sup>1</sup>, Michael F. Wells<sup>1</sup>, Talignair N. Venkatraman<sup>4</sup>, Christopher D. Lascola<sup>1,4</sup>, Zhanyan Fu<sup>1,5,6</sup> & Guoping Feng<sup>1,6,7</sup>

Autism spectrum disorders (ASDs) comprise a range of disorders that share a core of neurobehavioural deficits characterized by widespread abnormalities in social interactions, deficits in communication as well as restricted interests and repetitive behaviours. The neurological basis and circuitry mechanisms underlying these abnormal behaviours are poorly understood. SHANK3 is a postsynaptic protein, whose disruption at the genetic level is thought to be responsible for the development of 22q13 deletion syndrome (Phelan–McDermid syndrome) and other non-syndromic ASDs. Here we show that mice with *Shank3* gene deletions exhibit self-injurious repetitive grooming and deficits in social interaction. Cellular, electrophysiological and biochemical analyses uncovered defects at striatal synapses and cortico-striatal circuits in *Shank3* mutant mice. Our findings demonstrate a critical role for SHANK3 in the normal development of neuronal connectivity and establish causality between a disruption in the *Shank3* gene and the genesis of autistic-like behaviours in mice.

Autism and autism spectrum disorders (ASDs) are neurodevelopmental disorders diagnosed based on a triad of criteria: deficits in communication, impaired social interaction, and repetitive or restricted interests and behaviours<sup>1</sup>. ASDs are highly heritable disorders with concordance rates as high as 90% for monozygotic twins<sup>2</sup>. Recent genetic and genomic studies have identified a large number of candidate genes for ASDs<sup>3</sup>, many of which encode synaptic proteins<sup>4–6</sup>, indicating synaptic dysfunction may have a critical role in ASDs<sup>7,8</sup>. One of the most promising ASD candidate genes is *Shank3*, which codes for a key postsynaptic density (PSD) protein at glutamatergic synapses. Disruption of *Shank3* is thought to be the cause of core neurodevelopmental and neurobehavioural deficits in the 22q13 deletion syndrome (Phelan–McDermid syndrome), an autism spectrum disorder<sup>9–11</sup>. Furthermore, recent genetic screens have identified several mutations/rare variants of the *Shank3* gene in ASD patients outside of diagnosed 22q13 deletion syndrome<sup>12,13</sup>.

The Shank family of proteins (SHANK1–3) directly bind SAPAP (also known as DLGAP) to form the PSD-95–SAPAP–SHANK complex<sup>14,15</sup> (PSD-95 is also known as DLG4). This core of proteins is thought to function as a scaffold, orchestrating the assembly of the macromolecular postsynaptic signalling complex at glutamatergic synapses. Currently, however, little is known about the *in vivo* function of SHANK3 at the synapse and how a disruption of *Shank3* may contribute to ASDs. Here we demonstrate that genetic disruption of *Shank3* in mice leads to compulsive/repetitive behaviour and impaired social interaction, resembling two of the cardinal features of ASDs. Biochemical, morphological and electrophysiological studies revealed synaptic dysfunction at cortico-striatal synapses, part of the neural circuits strongly implicated as dysfunctional in ASDs. Our studies provide a synaptic and circuitry mechanism underlying *Shank3* disruption and ASD-like behaviours.

## *Shank3B*<sup>−/−</sup> mice display repetitive grooming

The *Shank3* gene codes for large proteins with multiple protein–protein interaction domains (Fig. 1a). We generated two different

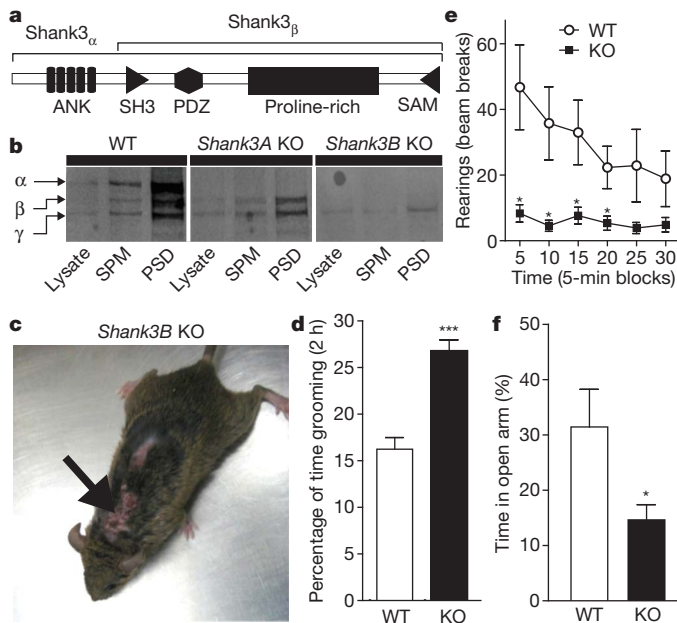
alleles of SHANK3 mutant mice. In *Shank3A* mutant mice, we targeted a portion of the gene encoding the ankyrin repeats (Supplementary Fig. 1b). This resulted in a complete elimination of SHANK3<sub>α</sub>, the longest SHANK3 isoform (Fig. 1b). However, the other two isoforms were not affected (here named SHANK3<sub>β</sub> and SHANK3<sub>γ</sub>). In *Shank3B* mutants, we targeted the fragment encoding the PDZ domain (Supplementary Fig. 1c). This led to the complete elimination of both SHANK3<sub>α</sub> and SHANK3<sub>β</sub> isoforms and a significant reduction of the putative SHANK3<sub>γ</sub> isoform at the PSD (−42.12% ± 9.27% of control, *n* = 3, *P* < 0.05) (Fig. 1b). Our analysis is mainly focused on the *Shank3B*<sup>−/−</sup> mutants due to their more pronounced behavioural and physiological defects.

We used mice with a hybrid genetic background to avoid the potential contribution to behavioural phenotypes of homozygous genetic variants on a pure inbred background<sup>7,16</sup>. Initially, F1 hybrids from heterozygous × heterozygous matings were generated and homozygous mice were born at an expected Mendelian rate. However, homozygous knockout mice from this type of mating are smaller than their wild-type littermates, presumably due to inadequate competition for resources during early postnatal days leading to different developmental trajectories. We postulated that this size difference would influence our behavioural tests. To alleviate this confound, heterozygous animals were crossed in direct brother-sister matings for five generations from which we derived F5 isogenic hybrids in a mixed background. These isogenic animals were then used to generate time-mated homozygous × homozygous breeding pairs to obtain wild-type and mutant animals used in the experiments. F5 *Shank3A* and F5 *Shank3B* knockouts from these matings are reared to weaning age with body weights similar to those from F5 control animals.

*Shank3B*<sup>−/−</sup> mice did not display any gross anatomical or histological brain abnormality, but on rare occasions exhibited seizures during handling in routine husbandry procedures. However, spontaneous seizures were never observed. By the age of 3–6 months, *Shank3B*<sup>−/−</sup>

<sup>1</sup>Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>2</sup>PhD Programme in Biomedicine and Experimental Biology (BEB), Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal. <sup>3</sup>Gulbenkian PhD Programme in Biomedicine, Gulbenkian Science Institute, 2781-901 Oeiras, Portugal. <sup>4</sup>Department of Radiology, and Brain Imaging and Analysis Center, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>5</sup>Department of Psychiatry and Behavioral Science, Duke University Medical Center, Durham, North Carolina 27710, USA. <sup>6</sup>McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>7</sup>Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts 02142, USA.

\*These authors contributed equally to this work.



**Figure 1 | Excessive grooming, skin lesions and anxiety-like behaviour in *Shank3B*<sup>-/-</sup> mice.** **a**, SHANK3 protein structure. **b**, Western blot showing a pan-SHANK3 antibody staining in brain lysate, synaptosomal plasma membrane (SPM) and 2× Triton X-100-washed PSD (PSD) fraction in wild-type (WT), *Shank3A*<sup>-/-</sup> and *Shank3B*<sup>-/-</sup> mice. **c**, Four-month-old *Shank3B*<sup>-/-</sup> mice display neck and head lesions (arrows). **d**, Pre-lesion *Shank3B*<sup>-/-</sup> (KO) mice spent more time in self-grooming than WT. **e**, In the open field test, *Shank3B*<sup>-/-</sup> mice, when compared to controls, display decreased rearing activity. **f**, In the zero maze test, *Shank3B*<sup>-/-</sup> mice spent less time in the open area than wild-type controls. \**P* < 0.05, \*\*\**P* < 0.001, two-tailed *t*-test for **d** and **f**, two-way repeated measures ANOVA with post hoc two-tailed *t*-test for **e**; all data are presented as means ± s.e.m. from 6–9 mice per genotype.

mice developed pronounced skin lesions with varying degrees of phenotypical penetrance: approximately 35% in the general holding colony (Fishers exact test, *P* < 0.0001), and 100% in mating females that have produced 4–6 litters. The lesions tend to appear first on the back of the neck or on the face (Fig. 1c) and usually progressed bilaterally to cover large areas of the body. The lesions were self-inflicted, as they were present in animals socially isolated at weaning age, and not due to excessive allogrooming, as no lesions were found in wild-type or *Shank3B*<sup>+/-</sup> mice housed from birth with *Shank3B*<sup>-/-</sup> animals. Furthermore, 24 h videotaping in pre-lesion animals revealed that *Shank3B*<sup>-/-</sup> mice showed an increase in time spent grooming when compared to wild-type controls (Fig. 1d). These observations indicate that *Shank3B*<sup>-/-</sup> mice display excessive grooming and self-injurious behaviour.

We characterized the animals further in a battery of behavioural tests. In the rotarod motor test, *Shank3B*<sup>-/-</sup> and control animals performed at similar levels (Supplementary Fig. 2). In the open field test, when compared to controls, *Shank3B*<sup>-/-</sup> mice showed similar levels of activity and thigmotaxis (Supplementary Fig. 2). However, rearing, which is a form of vertical exploration considered to be anxiogenic for mice, was significantly reduced in the mutants (Fig. 1e). In the elevated zero maze, the *Shank3B*<sup>-/-</sup> mice spent less time exploring the open arms of the maze versus the closed arms (Fig. 1f). In the light-dark emergence test, the *Shank3B*<sup>-/-</sup> mice displayed an increased latency to cross into the brightly lit area, although the time spent in each side of the box was similar between mutant animals and controls (Supplementary Fig. 2). Thus, the *Shank3B*<sup>-/-</sup> mice display an anxiety-like behaviour and excessive, self-injurious grooming. In contrast, *Shank3A*<sup>-/-</sup> mice displayed no lesions or anxiety-like behaviour (Supplementary Fig. 3).

## Social interaction deficits in *Shank3B*<sup>-/-</sup> mice

Deficits in social interaction are the most recognizable manifestation of autistic behaviours in humans. We used a modified version of a three-chamber social arena<sup>17</sup> to probe animals for their voluntary initiation of social interaction and their ability to discriminate social novelty. Initially, the test animal was left to explore and initiate social contact with a partner ('Stranger 1') held inside a wired cage or an identical but empty wired cage ('Empty cage'). In this test, the *Shank3B*<sup>-/-</sup> mice displayed dysfunctional social interaction behaviour, as measured by observing both time spent in the compartment containing the social partner (Fig. 2a, b) or in close interaction (Fig. 2d). Notably, *Shank3B*<sup>-/-</sup> mice exhibited a clear preference for interacting with the empty cage rather than with the social partner (Fig. 2a, d). In a subsequent trial, a novel social partner ('Stranger 2') was introduced into the previously empty wired cage. Wild-type mice displayed a preference for the novel animal, as shown by the increase in time spent in the compartment containing 'Stranger 2'. The *Shank3B*<sup>-/-</sup> mutants markedly spent more time in the centre chamber (Fig. 2c) and a reduced amount of time closely interacting with either social partner (Fig. 2e). In an identical test, the *Shank3A*<sup>-/-</sup> mice displayed normal initiation of social interaction, but perturbed recognition of social novelty (Supplementary Fig. 4).

Additionally, in an open arena test, freely interacting dyadic pairs of wild-type-*Shank3B*<sup>-/-</sup> mice displayed less time spent in reciprocal interaction, a lower frequency of nose-to-nose interaction and anogenital sniffing when compared to wild-type-wild-type pairs (Supplementary Fig. 5). Thus, data from both social interaction tests indicate that *Shank3B*<sup>-/-</sup> mice display abnormal social interaction as well as deficits in discriminating social novelty.

In our breeding scheme, *Shank3B*<sup>-/-</sup> mice and wild-type mice were nurtured by *Shank3B*<sup>-/-</sup> and wild-type dams, respectively. To assess the impact of maternal rearing on the observed sociability defects, we performed time-mated cross-fostering of *Shank3B*<sup>-/-</sup> mice and controls. Cross-fostering of *Shank3B*<sup>-/-</sup> neonatal pups with wild-type dams (KO<sub>cd</sub>) revealed qualitatively equivalent social defects in the mutant mice as compared to those observed in mutant mice nurtured by *Shank3B*<sup>-/-</sup> dams. Additionally, rearing wild-type neonatal pups by *Shank3B*<sup>-/-</sup> dams (WT<sub>cd</sub>) did not perturb normal sociability in wild-type animals (Supplementary Fig. 6). These data further indicate a genetic origin of the abnormal social behaviours in the *Shank3* mutant mice.

## Altered PSD composition in the striatum

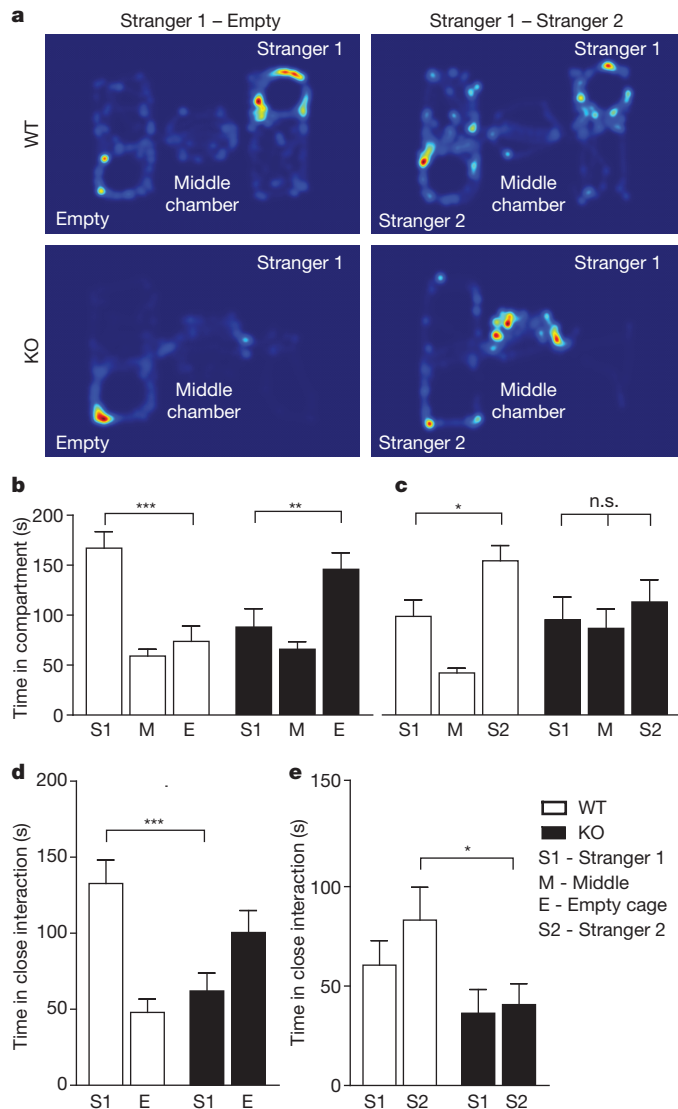
The basal ganglia are one of the brain regions implicated as dysfunctional in ASD. The repetitive/compulsive grooming behaviour in *Shank3B*<sup>-/-</sup> mice also suggests defects in cortico-striatal function. Furthermore, *Shank3*, but not *Shank1* or *Shank2*, is highly expressed in the striatum (Fig. 3a) (Supplementary Fig. 7 and Supplementary Table 1). Therefore, we focused our analyses on striatal neurons and cortico-striatal synapses.

Shank family members have been proposed as key regulators of the PSD at glutamatergic synapses<sup>18</sup>. To determine how the disruption of *Shank3* may affect the PSD protein network, we used biochemically purified PSDs from the striatum of wild-type and *Shank3B*<sup>-/-</sup> mice and performed semiquantitative western blotting for several scaffolding proteins (Fig. 3b) and glutamate receptor subunits (Fig. 3c). At the PSD level, we observed reduced levels of SAPAP3, Homer-1b/c and PSD-93 (also known as HOMER1 and DLG2, respectively; Fig. 3b) as well as a reduction in the glutamate receptor subunits GluR2, NR2A and NR2B (also known as GRIA2, GRIN2A and GRIN2B, respectively; Fig. 3c). These results suggest an altered molecular composition of postsynaptic machinery in the striatum and a possible disruption of glutamatergic signalling.

## Morphological defects of medium spiny neurons

To test whether disruption of SHANK3 affects neuronal morphology, we traced Golgi-stained striatal medium spiny neurons (MSNs) and

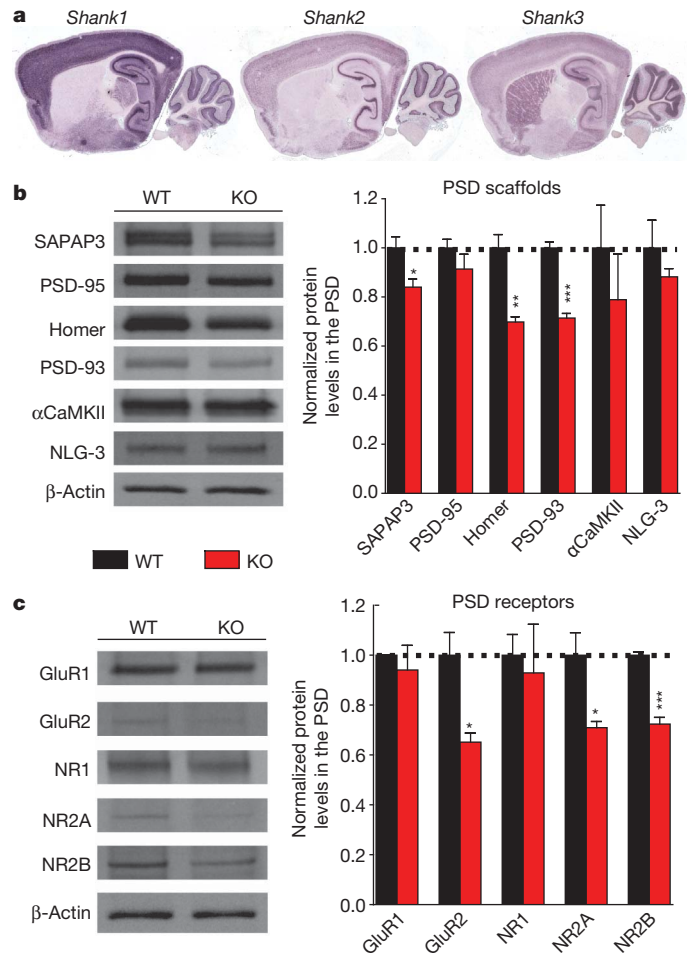




**Figure 2 | Reduced social interaction and abnormal social novelty recognition in *Shank3B*<sup>-/-</sup> mice.** **a**, Representative heat map analysis from 'Stranger 1 - Empty' and 'Stranger 1 - Stranger 2' trials from *Shank3B*<sup>-/-</sup> mice and controls. **b**, In the social interaction test, *Shank3B*<sup>-/-</sup> mice (closed bars) spent less time in the chamber containing the social partner (Stranger 1) and more time in the chamber containing the empty wire cage when compared to controls (open bars). **c**, In the social novelty test, *Shank3B*<sup>-/-</sup> mice do not display a preference for the novel social partner (Stranger 2), and spent more time in the middle chamber. **d**, **e**, When analysing social interaction by close proximity (within 5 cm) to either 'Stranger 1', 'Empty Cage' (**d**), or 'Stranger 1', 'Stranger 2' (**e**), *Shank3B*<sup>-/-</sup> mice displayed a clear reduction in social interaction when compared to controls (**d**); whereas under a social novelty paradigm (**e**), *Shank3B*<sup>-/-</sup> mice displayed a clear reduction in time spent with 'Stranger 2' \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.0001; one-way ANOVA, with Bonferroni post hoc *t*-test for **b**–**e**; all data presented as means ± s.e.m.; 12–14 mice per group.

their dendrites to investigate the cellular morphology and complexity of these cells. Sholl analysis revealed neuronal hypertrophy as measured by an increase in complexity of dendritic arborizations (Fig. 4a), total dendritic length (Fig. 4b) and also an increase in surface area (Fig. 4c) in *Shank3B*<sup>-/-</sup> MSNs.

Next, we performed patch-assisted Lucifer Yellow cell filling of MSNs and measured spine density in control and *Shank3B*<sup>-/-</sup> mice. *Shank3B*<sup>-/-</sup> mice displayed a significant reduction in spine density (Fig. 4d, e). We did not observe significant changes in spine length or head diameter; however, the neck width of *Shank3B*<sup>-/-</sup> MSN spines was slightly larger than that of controls (Supplementary Fig. 8).

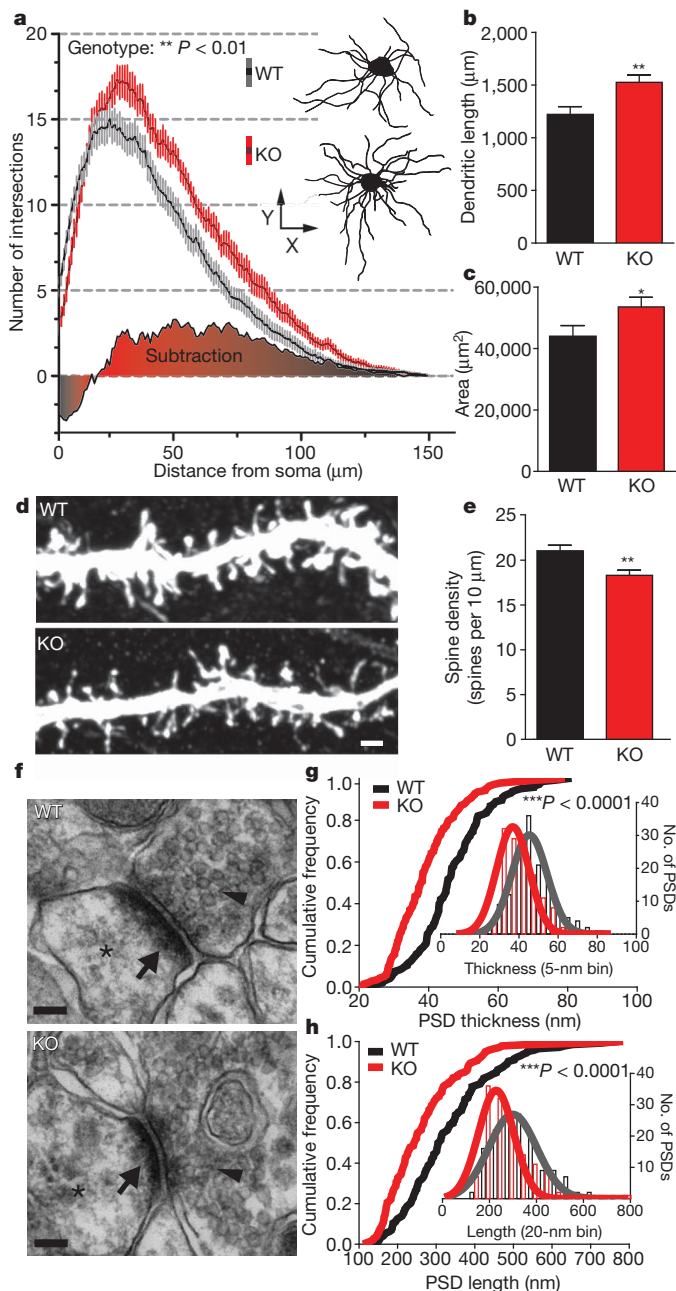


**Figure 3 | Biochemical changes in striatal synapses of *Shank3B*<sup>-/-</sup> mice.** **a**, Only *Shank3* mRNA is highly expressed in the striatum. **b**, Protein levels of the scaffolding proteins SAPAP3, Homer and PSD-93 are reduced in striatal PSD fractions from *Shank3B*<sup>-/-</sup> mice. αCaMKII and NLG-3 are also known as CAMK2A and NLGN3, respectively. **c**, Protein levels of glutamate receptor subunits GluR2, NR2A and NR2B are reduced in striatal PSD fractions from *Shank3B*<sup>-/-</sup> mice. GluR1 and NR1 are also known as GRIA1 and GRIN1, respectively. Each lane was loaded with 3 μg of protein with β-actin as loading control and normalized to wild-type levels. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, two-tailed *t*-test; all data are presented as means ± s.e.m.; *n* = 3 samples per group, with each sample being a combined pool of striatal tissue from three animals.

Finally we analysed PSD morphology by electron microscopy (Fig. 4f). We found a significant reduction in mean thickness (Fig. 4g) of PSDs from *Shank3B*<sup>-/-</sup> mice relative to controls. Additionally, PSD length was also significantly reduced in the *Shank3B*<sup>-/-</sup> mice (Fig. 4h). Taken together, these results highlight a critical *in vivo* role for *Shank3* in the normal development of medium spiny neurons and striatal glutamatergic synapses.

### Striatal hypertrophy in *Shank3B*<sup>-/-</sup> mice

Even though there is no clear correlation between brain size or neuronal hypertrophy specifically for *Shank3* disruptions in humans, a potential link between enlarged brain size, neuronal hypertrophy and autism has been suggested previously<sup>19</sup>. In particular, increased caudate volume in autism patients has been proposed to be linked to repetitive behaviours<sup>20,21</sup>. We measured striatal volume using three-dimensional magnetic resonance imaging in the intact brain of *Shank3B*<sup>-/-</sup> and control mice. We found that there was no significant difference in overall brain size between the genotypes. However, measurement of caudate volume in the same animals revealed a small but

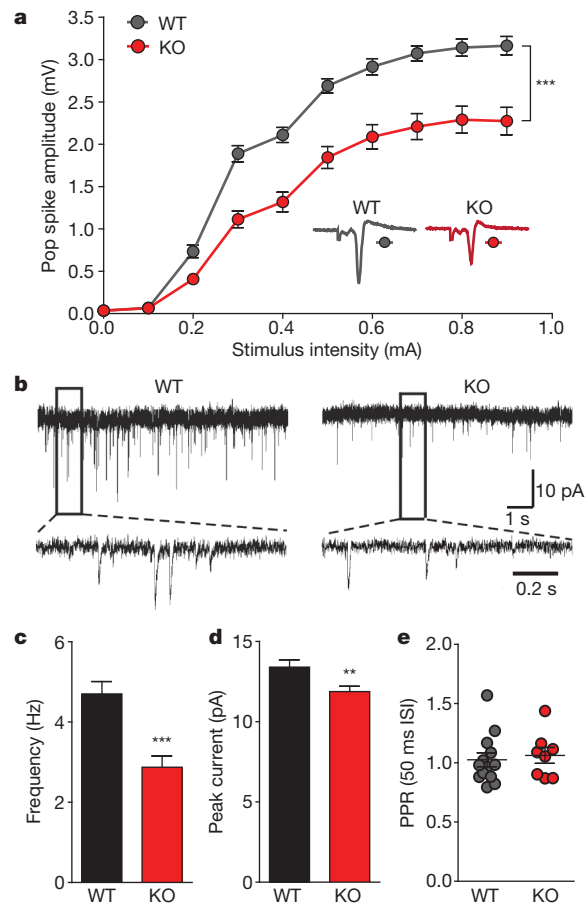


**Figure 4 | Morphological and ultrastructural neuronal abnormalities in *Shank3B*<sup>-/-</sup> mice.** **a**, Sholl analysis reveals an increased neuronal complexity of *Shank3B*<sup>-/-</sup> MSNs (red) when compared to MSNs from wild-type mice (grey); example neurons are shown as insets (top, WT-; bottom, KO-). **b**, **c**, MSNs from *Shank3B*<sup>-/-</sup> mice show an increase in total dendritic length (**b**) and surface area (**c**) when compared to controls. **d**, Representative confocal stacks of dye-filled MSNs from KO and WT mice; scale bar, 1  $\mu\text{m}$ . **e**, Spine density in MSNs from *Shank3B*<sup>-/-</sup> mice is lower than that of wild-type MSNs. **f**, Examples of electron micrographs depicting the synaptic contacts with presynaptic vesicles (arrowheads), postsynaptic densities (arrow) and dendritic spine (asterisk); scale bar, 100 nm. **g**, *Shank3B*<sup>-/-</sup> PSDs are thinner than wild-type PSDs. **h**, *Shank3B*<sup>-/-</sup> PSDs are shorter than wild-type PSDs. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.0001$ ; two-way repeated measures ANOVA for **a**; two-tailed  $t$ -test for **b**, **c** and **e**; two-sample Kolmogorov-Smirnov test for **g** and **h**. Data in **g** and **h** are presented as cumulative frequency plot with histogram distribution and Gaussian curve fit for the insets. Data from **b**, **c** and **e** are presented as means  $\pm$  s.e.m.;  $n = 36$  from 3 wild-type mice and  $n = 36$  from 3 *Shank3B*<sup>-/-</sup> mice for **a**-**c**;  $n = 41$  dendritic segments from 3 wild-type mice and  $n = 36$  dendritic segments from 3 *Shank3B*<sup>-/-</sup> mice for **e**;  $n = 144$  PSDs from three wild-type mice and  $n = 140$  PSDs from three *Shank3B*<sup>-/-</sup> mice for **g**, **h**.

significant volumetric enlargement of this structure in *Shank3B*<sup>-/-</sup> mice (Supplementary Fig. 9). These data suggest a correlation between neuronal hypertrophy and brain volume, consistent with studies from other mouse models of ASD<sup>22,23</sup>.

### Perturbation of striatal postsynaptic function

To elucidate the functional consequences of a disruption in *Shank3* on synaptic function, we performed recordings of cortico-striatal synaptic circuitry in acute brain slices of 6–7-week-old animals. We found that field population spikes were significantly reduced in *Shank3B*<sup>-/-</sup> mice when compared with controls (Fig. 5a). Presynaptic function was not altered, as indicated by the relationship of stimulation intensity to the amplitude of the action potential component of the response termed negative peak 1 (NP1) and the paired-pulse ratio (PPR; Supplementary Fig. 10). These results indicate that the reduction in total field responses was most likely due to a postsynaptic impairment in synaptic function and/or a reduction in the number of functional synapses. Consistent with their mild behavioural phenotypes *Shank3A*<sup>-/-</sup> mice displayed minimal disruption at cortico-striatal synapses (Supplementary Fig. 11).



**Figure 5 | Reduced cortico-striatal synaptic transmission in *Shank3B*<sup>-/-</sup> MSNs.** **a**, Cortico-striatal pop spike amplitude is decreased in *Shank3B*<sup>-/-</sup> mice (red trace) as measured by extracellular field recordings. Inset, example traces for *Shank3B*<sup>-/-</sup> (KO) and wild-type (WT). **b**, mEPSC example traces from wild-type and *Shank3B*<sup>-/-</sup> MSNs recorded with whole-cell voltage clamp. **c**, **d**, Reduced mEPSC frequency (**c**) and amplitude (**d**) in *Shank3B*<sup>-/-</sup> MSNs when compared to wild-type. **e**, PPR is unaltered in *Shank3B*<sup>-/-</sup> MSNs. \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ; two-way repeated measures ANOVA, with Bonferroni post hoc test for **a**; two-tailed  $t$ -test for **c**, **d**; all data presented as means  $\pm$  s.e.m. For field recordings,  $n = 13$  slices from four mice per group; for mEPSCs,  $n = 29$  MSNs from wild-type mice,  $n = 32$  MSNs from *Shank3B*<sup>-/-</sup> mice.

We next performed whole-cell voltage clamp recordings of  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor-miniature excitatory postsynaptic currents (AMPA-mEPSCs) in dorsolateral striatal MSNs. We found that the frequency of mEPSCs was significantly reduced in *Shank3B*<sup>-/-</sup> MSNs (Fig. 5b, c), indicating a reduction in the number of functional synapses in *Shank3B*<sup>-/-</sup> MSNs because we did not observe defects on presynaptic function by measuring PPR (Fig. 5e). We also found a significant reduction of peak mEPSC amplitude in *Shank3B*<sup>-/-</sup> MSNs (Fig. 5b, d), indicating a reduction in the postsynaptic response from the available synapses. We did not observe significant differences in *N*-methyl-D-aspartate receptor (NMDA)/AMPA receptor-mediated current ratio in *Shank3B*<sup>-/-</sup> neurons (Supplementary Fig. 12). Finally, similar defects in mEPSC frequency and amplitude were observed in *Shank3B*<sup>-/-</sup> and wild-type littermate mice obtained from heterozygous matings (Supplementary Fig. 13). Together, these data demonstrate a critical role for SHANK3 in postsynaptic function in cortico-striatal circuitry.

To assess if the defects arising from *Shank3* dysfunction were specific to striatal circuitry or due to a more broad CNS perturbation, we performed a Morris water maze task for hippocampal-dependent learning and memory. We found that *Shank3B*<sup>-/-</sup> mice performed at the same levels as controls in both learning and probe trials (Supplementary Fig. 14a–c). Reversal learning and probe trials again demonstrated similar levels of performance between *Shank3B*<sup>-/-</sup> mice and controls (Supplementary Fig. 14d–f). Concomitantly, we performed electrophysiological recordings from the hippocampal CA1 sub-region and found no obvious difference in field recordings of population spikes or PPR between genotypes (Supplementary Fig. 15a–c). In addition, we found no significant differences in mEPSC frequency or mEPSC amplitude (Supplementary Fig. 15d–f). These data suggest that the observed behavioural and synaptic defects are specific to discrete brain regions and are not part of an overall CNS dysfunction.

## Discussion

Despite recent advances in the understanding of autism spectrum disorder genetics, the underlying neurobiological substrates and neural circuits involved in these disorders remain largely unknown. The *Shank3* gene has become the focus of substantial interest, with an increasing body of evidences suggesting *Shank3* as the causative gene of the major neurological symptoms in the 22q13 deletion syndrome<sup>9,11–13,24</sup>. Our present study with *Shank3* mutant mice not only sheds light on a critical *in vivo* role for SHANK3 in striatal glutamatergic synaptic structure and function, but also demonstrates causality between a disruption in this gene and the development of autistic-like behaviours in mice.

In this study, we generated two mutant alleles for the *Shank3* gene. These two lines of mice showed different levels of severity in synaptic defects and phenotypes. In humans, multiple mutations/variants of *Shank3* gene have been identified to coalesce at the ankyrin repeats and downstream of PDZ domain<sup>9,13</sup>. Our data indicate that disruptions of different locations of the *Shank3* gene can lead to varying degrees of functional defects, which may in part contribute to phenotypic heterogeneity in *Shank3*-related ASDs. We should note that, in clinical conditions, the 22q13 deletions and the autism-associated *Shank3* mutations are heterozygous, whereas in our current study, we used homozygous mutant mice to get a clear understanding of the physiological role of the *Shank3* gene and the underlying functional consequences of its disruption. Further studies will be needed to elucidate potential functional deficits resulting from *Shank3* haploinsufficiency in *Shank3B*<sup>-/-</sup> mice.

PSD-95–SAPAP–SHANK proteins form a key postsynaptic scaffold at glutamatergic synapses which interacts with many synaptic proteins, including the neuroligin–neurexin complex<sup>25</sup>. In addition to *Shank3* (ref. 9), it is worth noting that *Shank2* (refs 6, 8), *SAPAP2* (ref. 6), neuroligin-1 (ref. 26) and neuroligin-3 and -4 (ref. 27) have all been

implicated in human ASDs. Therefore, the dysfunction of neuroligin–neurologlin–PSD-95–SAPAP–SHANK complex could underlie a common synaptic mechanism for a subset of ASDs.

The precise circuitry defects involved in autistic behaviours are poorly understood. Neuroimaging studies provide evidence that caudate and frontal-striatal circuitries are dysfunctional areas in ASD<sup>28–30</sup>. Cortico-striatal circuitry dysfunction has also been strongly implicated in repetitive/compulsive behaviours in obsessive-compulsive disorder (OCD)<sup>31–33</sup>. We previously found that deletion of *SAPAP3*, which directly interacts with SHANK3 and is highly expressed in the striatum, leads to cortico-striatal circuitry dysfunction and OCD-like behaviours including repetitive/compulsive grooming in mice<sup>31</sup>. Repetitive behaviours are also often seen in autistic patients and in some mouse models of ASDs<sup>34–36</sup>. SHANK3 is the most abundant SHANK family member expressed in the striatum and *Shank3B*<sup>-/-</sup> mice exhibit excessive/repetitive grooming leading to skin lesions. Our data support the hypothesis that repetitive behaviours in OCD and ASD may share a common circuitry mechanism.

The regulation of social behaviours and social interaction is thought to be controlled by several brain regions and circuits<sup>37</sup>. Similarly, genetic makeup is thought to have a key role in the phenotypic manifestation of social behaviours<sup>38</sup>. The robust social interaction deficits in *Shank3B* mutant mice demonstrate a casual role for the disruption of this gene in the genesis of social dysfunction and provide a valuable experimental system for future genetic dissection of the neuronal basis of social behaviour.

## METHODS SUMMARY

**Behavioural analysis.** Young adult mice 5–6-weeks old were used for all behavioural analyses except lesion scores which were performed in 4–5-month-old mice. All experiments were done blind to genotypes. All experimental procedures were reviewed and approved by the Duke University Institutional Animal Care and Use Committee and the MIT Committee on Animal Care.

**Statistical analysis.** Analyses were performed using Prism (GraphPad Software) and MATLAB (MathWorks). Details on particular tests used are described in the main text and in the methods section; a summary of statistical analysis for the behavioural data are presented in Supplementary Table 2.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 7 September 2010; accepted 22 February 2011.

Published online 20 March 2011.

1. American Psychiatric Association Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (American Psychiatric Association, 2000).
2. Rosenberg, R. E. *et al.* Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163**, 907–914 (2009).
3. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9**, 341–355 (2008).
4. Bourgeron, T. A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19**, 231–234 (2009).
5. Zoghbi, H. Y. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* **302**, 826–830 (2003).
6. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
7. Tabuchi, K. *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318**, 71–76 (2007).
8. Berkel, S. *et al.* Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genet.* **42**, 489–491 (2010).
9. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39**, 25–27 (2006).
10. Prasad, C. *et al.* Genetic evaluation of pervasive developmental disorders: the terminal 22q13 deletion syndrome may represent a recognizable phenotype. *Clin. Genet.* **57**, 103–109 (2000).
11. Wilson, H. L. *et al.* Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* **40**, 575–584 (2003).
12. Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81**, 1289–1297 (2007).
13. Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **150B**, 421–424 (2009).
14. Kim, E. *et al.* GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136**, 669–678 (1997).



15. Takeuchi, M. *et al.* SAPAPs. A family of PSD-95/SAP90-associated proteins localized at postsynaptic density. *J. Biol. Chem.* **272**, 11943–11951 (1997).
16. Zoghbi, H. Y. & Warren, S. T. Neurogenetics: advancing the “next-generation” of brain research. *Neuron* **68**, 165–173 (2010).
17. Moy, S. S. *et al.* Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav.* **3**, 287–302 (2004).
18. Hung, A. Y. *et al.* Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. *J. Neurosci.* **28**, 1697–1708 (2008).
19. Redcay, E. & Courchesne, E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol. Psychiatry* **58**, 1–9 (2005).
20. Langen, M. *et al.* Changes in the developmental trajectories of striatum in autism. *Biol. Psychiatry* **66**, 327–333 (2009).
21. Hollander, E. *et al.* Striatal volume on magnetic resonance imaging and repetitive behaviors in autism. *Biol. Psychiatry* **58**, 226–232 (2005).
22. Bourgeron, T. A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19**, 231–234 (2009).
23. Kwon, C. H. *et al.* Pten regulates neuronal arborization and social interaction in mice. *Neuron* **50**, 377–388 (2006).
24. Bonaglia, M. C. *et al.* Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J. Med. Genet.* **43**, 822–828 (2006).
25. Irie, M. *et al.* Binding of neuroligins to PSD-95. *Science* **277**, 1511–1515 (1997).
26. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 199–207 (2008).
27. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34**, 27–29 (2003).
28. Silk, T. J. *et al.* Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: a functional MRI study. *Am. J. Psychiatry* **163**, 1440–1443 (2006).
29. Horwitz, B., Rumsey, J. M., Grady, C. L. & Rapoport, S. I. The cerebral metabolic landscape in autism. Interrelations of regional glucose utilization. *Arch. Neurol.* **45**, 749–755 (1988).
30. Sears, L. L. *et al.* An MRI study of the basal ganglia in autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **23**, 613–624 (1999).
31. Welch, J. M. *et al.* Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. *Nature* **448**, 894–900 (2007).
32. Shmelkov, S. V. *et al.* Slitrk5 deficiency impairs corticostriatal circuitry and leads to obsessive-compulsive-like behaviors in mice. *Nature Med.* **16**, 598–602 (2010).
33. Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31**, 359–387 (2008).
34. McFarlane, H. G. *et al.* Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav.* **7**, 152–163 (2008).
35. Blundell, J. *et al.* Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J. Neurosci.* **30**, 2115–2129 (2010).
36. Etherton, M. R., Blaiss, C. A., Powell, C. M. & Sudhof, T. C. Mouse neuroligin-1 $\alpha$  deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proc. Natl Acad. Sci. USA* **106**, 17998–18003 (2009).
37. Insel, T. R. & Fernald, R. D. How the brain processes social information: searching for the social brain. *Annu. Rev. Neurosci.* **27**, 697–722 (2004).
38. Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S. & Knafo, A. Genetics of human social behavior. *Neuron* **65**, 831–844 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank C. Duarte, S. Chatterjee and A. Oliveira-Maia for discussions; L. Kruger and Q. Liu for technical assistance; A. Hadiono for assistance in behavioural annotation; D. Bredt for the PSD-93 antibody; T. Boeckers for the anti-SHANK3 antibody; S. Miller and P. Christopher for advice and assistance with electron microscopy techniques; J. Crawley for the demonstration of social behaviour tests; N. Calakos and Y. Wan for advice on electrophysiology studies; A. Graybiel for critical comments of the manuscript; D. Wang and the other members of the G.F. laboratory for their support. We thank The Poitras Center for Affective Disorders Research. This work was funded by a grant from NIMH/NIH (R01MH081201), a Hartwell Individual Biomedical Research Award from The Hartwell Foundation, and a Simons Foundation Autism Research Initiative (SFARI) grant Award to G.F.; a NARSAD Young Investigator Award and NIH Ruth L. Kirschstein National Research Service Award (F32MH084460) to J.T.T.; a NIH (R03MH085224) grant to Z.F.; and doctoral fellowships from the Portuguese Foundation for Science and Technology to J.P. (SFRH/BD/15231/2004) and C.F. (SFRH/BD/15855/2005). C.F. would like to acknowledge the support from the “Programa Gulbenkian de Doutoramento em Biomedicina” (PGDB, Oeiras, Portugal) and J.P. the “Programa Doutoral em Biologia Experimental e Biomedicina” (CNC, Coimbra, Portugal).

**Author Contributions** J.P., C.F., J.T.T., W.W., M.F.W., T.N.V., C.D.L. and Z.F. participated in the execution and analysis of experiments. J.P., C.F., J.T.T., C.D.L., Z.F. and G.F. participated in the interpretation of the results. J.P., C.F. and G.F. designed the experiments and wrote the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to G.F. (fengg@mit.edu).

## METHODS

**Mice.** *Shank3* mutant mice were generated by homologous recombination in R1 embryonic stem cells and implanted in C57 blastocysts using standard procedures. One targeting vector (*Shank3A*) was designed to replace exon 4–7 (containing the ankyrin repeat domains) and another vector (*Shank3B*) was designed to replace exon 13–16 (containing the PDZ domain) of the *Shank3* gene with a NEO cassette. Genotypes were determined by PCR of mouse tail DNA, using: for *Shank3A*, primer F1a (GGTTGAGGATGAGCAAGCTAG) and R1a (GGGACATAAGTGAAGGTTAGG) for the wild-type allele (318 base pairs), and F1a and R2 (TCAGGGTTATTGTCTCATGAGC; in the neo cassette) for the mutant allele (361 base pairs); for *Shank3B*, primer F1b (GAGCTCTACTCCCTT AGGACTT) and R1b (TCCCCCTTCACTGGACACCC) for the wild-type allele (316 base pairs), and F1b and R2 (TCAGGGTTATTGTCTCATGAGC; in the neo cassette) for the mutant allele (360 base pairs). The NEO cassette was not removed.

Chimaeric mice were crossed to C57 females (Jackson Labs). Initially, F1 hybrids from heterozygous  $\times$  heterozygous matings were generated. However, homozygous knockouts mice from this type of mating are smaller than their wild-type littermates, presumably due to an inadequate competition for resources during early postnatal days leading to different developmental trajectories. We postulated that this size difference would influence our behavioural tests. To alleviate this confound, heterozygous animals were crossed in direct brother-sister matings for five generations from which we derived F5 isogenic hybrids in a mixed background. These isogenic animals were then used to generate time-mated homozygous  $\times$  homozygous breeding pairs to obtain wild-type and mutant animals used in the experiments. F5 *Shank3A* and F5 *Shank3B* knockouts from these matings are reared to weaning age with weights similar to those from F5 control animals.

Animals were housed at a constant 23 °C in a 12 h light/dark cycle (lights off at 19:00), with food and water available *ad libitum*. Mice were housed 3–5 by genotype per cage with the exception of the animals individually housed for grooming measurements. Only aged-matched male mice were used for behavioural experiments, all other tests included age-matched males and females in proportional contribution across groups. Unless otherwise noted, all tests were conducted with naive cohorts of mice. All experimental procedures were reviewed and approved by the Duke University Institutional Animal Care and Use Committee and the MIT Committee on Animal Care.

**Grooming behaviour<sup>31</sup>.** Young adult male mice 5–6-week-old were used for analysis of grooming behaviour. Habituated, individually housed animals were video-taped for 24 h under 700 lx (day, 12 h) and  $\sim$ 2 lx (red light at night, 12 h) illumination. Grooming behaviours were coded from 19:00–21:00 h (that is, 2 h beginning at the initiation of the dark cycle); this segment was analysed using Noldus Observer software and the total amount of time in the 2-h segment spent grooming was determined. Grooming included all sequences of face-wiping, scratching/rubbing of head and ears, and full-body grooming. The observer was blinded to genotype during the scoring of the videotapes.

**PSD preparation and western blot.** PSD fractions of the striatum were prepared as previously described<sup>31</sup>, separated on SDS–PAGE and probed with specific antibodies. The relative amount of  $\beta$ -actin was used as loading control. Antibodies used in these experiments include rabbit antibodies against PSD-93 (gift from D. Bredt) and *Shank3* (gift from T. Boeckers). The antibody for SAPAP3 has been previously described<sup>39</sup>. Commercial antibodies used include monoclonal antibodies against NR1 (Transduction Laboratories), NR2B (Millipore), CaMKII (Transduction Laboratories), NR2A (Millipore), and  $\beta$ -actin (Sigma), as well as polyclonal antibodies against GluR1 (Abcam), Homer (Chemicon), GluR2 (Abcam), neuroligin-3 (Synaptic Systems) and PSD-95 (Abcam).

**In situ hybridization.** mRNA *in situ* hybridization was performed as described elsewhere<sup>39</sup>. Briefly, reactions were performed with 20  $\mu$ m cryosections from freshly frozen 5-week-old brain mouse tissue using digoxigenin (DIG)-labelled riboprobes against mouse *Shank1* cDNA (NM\_001034115; base pairs 4107–4924), *Shank2* cDNA (NM\_001081370; base pairs 2063–2876) and *Shank3* (NM\_021423; base pairs 3159–3959). The complementary DNAs used were all verified by sequencing compared to the following sequences GenBank accession numbers: (*Shank1*: NM\_001034115), (*Shank2*: NM\_001081370) and (*Shank3*: NM\_021423). The hybridization signal was detected using an alkaline phosphatase (AP)-conjugated anti-DIG antibody (Roche) and developed using 5-bromo-4-chloro-indolylphosphate/nitroblue tetrazolium (BCIP/NBT; Roche).

**Motor and anxiety-like behaviours<sup>31</sup>.** Zero maze: an elevated zero maze was indirectly illuminated at 100 lx. Testing commenced with an animal being introduced into a closed area of the maze. Behaviour was video-taped for 5 min and subsequently scored by a trained observer using Noldus Observer software. Anxiety-like behaviour was deduced based upon the percent time spent in the open areas. The observer was blinded to genotype. The animals used in the zero

maze test, both *Shank3A*<sup>−/−</sup>, *Shank3B*<sup>−/−</sup> and respective controls were previously tested in the open field test with a 2-day period in between tasks.

**Open field:** spontaneous locomotor activity was evaluated over 30 min in an automated Omnitech Digiscan apparatus (AccuScan Instruments) as described<sup>31</sup>. Locomotor activity was assessed as total distance travelled (m). Anxiety-like behaviour was defined by number of rearings and time spent in the centre as compared to time spent in the perimeter (thigmotaxis) of the open field.

**Dark-light emergence test:** mice were habituated in an adjacent room to low light conditions ( $\sim$ 40 lx) and the test room was initially under similar illumination. Testing was conducted in a two-chambered test apparatus (Med Associates), with one side draped in black cloth (that is, dark-chamber) and the other illuminated at  $\sim$ 1,000 lx (that is, light-chamber) with a high intensity house light and overhead fluorescent lamps. Upon placing the mice into the dark chamber, the light chamber was illuminated and the door between the two chambers was opened. The mice were allowed to freely explore the apparatus for 5 min. The latency to emerge from the darkened into the lighted chamber and the percentage of time spent in the illuminated chamber were used as indices of anxiety-like behaviours.

**Social interaction paradigm.** Three-chamber social test: sociability and response to social novelty test was performed as previously described<sup>17</sup> with minor modifications. Briefly, 5–6-week-old male animals were used across all tests. Target subjects (Stranger 1 and Stranger 2) were 5–6-week-old males habituated to being placed inside wire cages for 5 days before beginning of testing. Test mice were habituated to the testing room for at least 45 min before the start of behavioural tasks. The social test apparatus consisted of a transparent acrylic box with removable floor and partitions dividing the box into three chambers. Here, the middle chamber (20 cm  $\times$  17.5 cm) is half the width of Chamber 1 (20 cm  $\times$  35 cm) and Chamber 2 (20 cm  $\times$  35 cm) with the overall dimensions of the box being 60 cm (length)  $\times$  35 cm (width) with 5 cm openings between each chamber which can be closed or open with a lever operated door. The wire cages used to contain the stranger mice were cylindrical, 11 cm in height, a bottom diameter of 10.5 cm with the bars spaced 1 cm apart (Galaxy Cup, Spectrum Diversified Designs). An inverted transparent cup was placed on the top of the cage to prevent the test mice from climbing on the top of the wire cage.

For the sociability test, the test animal was introduced to the middle chamber and left to habituate for 5 min, after which an unfamiliar mouse (Stranger 1) is introduced into a wire cage in one of the side-chambers and an empty wire cage on the other side-chamber. The dividers are then raised and the test animal is allowed to freely explore all three chambers over a 5 min session. Following the 5 min session, the animal remains in the chamber for an extra 5 min (post-test) to better acquire the identification cues from Stranger 1 animal. Following this, a novel stranger mouse (Stranger 2) is inserted in the wire cage previously empty and again the test animal is left to explore for a 5 min session. Time spent in each chamber, time spent in close proximity and heat maps were calculated using the automated software Noldus Ethovision. The release of the animals and relative position of social and inanimate targets was counterbalanced. However, for each individual test animal the location of Stranger 1 was maintained during Stranger 1 – E and Stranger 1 – Stranger 2 testing of the social behaviour.

**Dyadic social interaction:** animals were acclimatized to the test room for at least 1 h before the experiment. Target mice were wild-type and *Shank3B*<sup>−/−</sup> of 6 weeks of age. Stimulus mice were conspecific age-matched wild-type mice socially naive to the target mice. At least 3 h before the beginning of the test, stimulus mice were given identifiable markings on the tails using a black marker pen. A pair of target and stimulus mice were introduced in a transparent Plexiglas arena (40 cm  $\times$  40 cm  $\times$  30 cm) covered with fresh bedding and the session recorded for 10 min. Quantification of social behaviours was performed using Noldus Observer software by a researcher blinded to the genotype of the target animals. Quantifications included: reciprocal social interaction, as determined by any sequence or combination of sequences involving close huddling, sniffing (for example, nose-to-nose, anogenital sniffing) or allogrooming by the target and stimulus mouse; the frequency of nose-to-nose sniffing; and the frequency of anogenital sniffing initiated by the target animal towards the stimulus mouse. Statistical tests were performed using unpaired two-tailed *t*-test.

**Rotarod.** Motor coordination was assessed in an accelerating rotarod test (4–40 r.p.m.). Briefly, animals were introduced in the apparatus (Med Associates) and the latency to fall was determined. Animals were tested for three trials in a single day with an inter-trial interval of 10 min.

**Morris water maze.** Morris water maze testing was conducted as describe elsewhere<sup>40</sup> with minor modifications. Male mice (4–5-weeks old) selected for the test were individually handled daily for 5 days before beginning the experiment. Testing pool was 120 cm in diameter and the platform 8 cm in diameter. The platform was submerged 1 cm below the water surface. Pool water was maintained at 23.0  $\pm$  0.5 °C and made opaque by mixing-in white non-toxic tempera paint. During training, 90 s duration trials were used, if the animals did not find

the platform within 90 s the experimenter guided the animal to the platform. After reaching the platform the animals were left for 15 s on top of the platform before being removed. Trials were administered for 5 days with four trials per animal per day with the platform located in the south-west quadrant. On the sixth day a 60 s probe trial was performed. On the seventh day, the reversal training commenced with the platform in the north-east quadrant, and proceeded as described above. The experimenter followed the animals' progress using tracking software outside of the testing room. Tracking and analysis were performed using the Noldus Ethovision software.

**Golgi staining and Sholl analysis.** All brains and collected sections were coded in order to blind the experimenter of the genotype until after all data was collected and analysed. Brains from 5-week-old, gender-matched *Shank3B*<sup>-/-</sup> and control mice were prepared using standard Golgi–Cox impregnation technique using the FD Rapid GolgiStain Kit (NeuroTechnologies). Serial coronal sections of 100  $\mu$ m were collected from controls and *Shank3B* mutant animals. A total of 12 cells per animal were traced across the dorsal striatum as to sample representatively from this structure for a final number of 36 cells per genotype. For each animal, sections were selected to be between rostral-caudal bregma 1.18 mm and 0.86 mm. Criteria to identify medium spiny neurons were, (1) presence within the caudate putamen; (2) full impregnation of the neuron along the entire length of the dendritic arborization; (3) relative non-overlap with surrounding neurons and isolation from astrocytes and blood vessels and (4) morphologically, by the presence of high number of spines and relatively short neuronal arborizations as characteristics of MSNs. For each selected neuron the entire neuronal arbor was reconstructed under a  $\times 100$  oil lens in a motorized microscope with a digital CCD camera connected to a computer running Neurolucida Software (MBF Bioscience). The three-dimensional analysis of the reconstructed neurons was performed using NeuroExplorer software (MBF Bioscience) and data from branch length, number of branches and neuronal complexity was measured and analysed in Prism (Graph Pad). Two-way repeated measures ANOVA was used for Sholl analysis. Statistical significance was accepted when  $*P < 0.05$ ,  $**P < 0.01$  and  $***P < 0.0001$ .

**Cortico-striatal electrophysiology.** Brain slice preparation for extracellular field recording: acute brain slices were prepared from 6–7-week-old mice. Slices were prepared from one WT and one KO pair each day and the experimenter was blinded to the genotype. The mice were deeply anesthetized by intra-peritoneal injection of avertin and then transcardially perfused with carbogenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>) ice-cold protective cutting artificial cerebrospinal fluid (aCSF) with the composition (in mM): 119 glycerol, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 25 glucose, 2 thiourea, 5 L-ascorbic acid, 3 Na-pyruvate, 0.5 CaCl<sub>2</sub>·4H<sub>2</sub>O, 10 MgSO<sub>4</sub>·7H<sub>2</sub>O. Mice were then decapitated and the brains were removed into ice-cold cutting solution for an additional 1 min. The brains were then rapidly blocked for coronal sectioning at 300- $\mu$ m thickness on a VF200 model compresstome (Precisionary Instruments) using either a sapphire or zirconium ceramic injector style blade. Slices containing the dorsal striatum were initially recovered for 30 min at room temperature (23–25 °C) in a carbogenated protective recovery aCSF (same composition as the cutting aCSF except that glycerol was replaced with *N*-methyl-D-glucamine (NMDG)-Cl as a substitute for NaCl to prevent initial excitotoxic swelling during re-warming). After this initial 30 min period the slices were transferred into a holding chamber containing carbogenated normal aCSF of the composition (in mM): 119 NaCl, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 12.5 glucose, 2 CaCl<sub>2</sub>·4H<sub>2</sub>O, 1 MgSO<sub>4</sub>·7H<sub>2</sub>O. The holding aCSF was supplemented with (in mM): 2 thiourea, 1 L-ascorbic acid, 3 Na-pyruvate to improve slice health and longevity, and slices were stored for 1–6 h before transfer to the recording chamber for use. The osmolality of all solutions was measured at 300–310 mOsm and the pH was maintained at  $\sim 7.3$  after equilibration under constant carbogenation.

Supplementary Fig. 11 shows summary data for corticostriatal field recordings from acute coronal brain slices of *Shank3A* mutant versus WT mice. The method of slice preparation differed significantly in these earlier experiments. Mice were transcardially perfused with carbogenated ice-cold protective sucrose aCSF with the composition (in mM): 185 sucrose, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 25 glucose, 0.5 CaCl<sub>2</sub>·4H<sub>2</sub>O, 4 MgSO<sub>4</sub>·7H<sub>2</sub>O (pH 7.3, 300–310 mOsm) without supplementation of antioxidants. Slices were immediately transferred into a holding chamber containing carbogenated normal aCSF of the composition (in mM): 119 NaCl, 2.5 KCl, 1.25 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 12.5 glucose, 2 CaCl<sub>2</sub>·4H<sub>2</sub>O, 1 MgSO<sub>4</sub>·7H<sub>2</sub>O (pH 7.3, 300–310 mOsm) without supplementation of antioxidants, and slices were stored for 1–4 h before transfer to the recording chamber. The absence of the initial 30 min recovery period in 'protective' aCSF in addition to the absence of antioxidant supplementation in the cutting aCSF and in the aCSF in the holding chamber results in more rapid deterioration of slice health and smaller evoked population spike amplitudes on average, indicating reduced overall slice viability compared to slices prepared with a 30 min NMDG aCSF recovery protocol described above. However, WT and KO brain slices were

always subjected to identical procedures on any given day of recording and the procedures were always standardized for each discrete experimental data set so that these factors would not introduce any potential confounds.

**Extracellular field recording.** A platinum iridium concentric bipolar stimulating electrode (CBAPC75, 25  $\mu$ m inner pole diameter; FHC) was placed on the inner border of the corpus callosum between the cortex and dorsolateral striatum. This electrode position was chosen to predominantly activate cortical axons within the corpus callosum which heavily converge upon striatal MSNs to form excitatory corticostriatal synaptic connections. Although there is ample evidence on which to base our assertion that stimulation of the corpus callosum predominantly results in activation of cortical axons<sup>41,42</sup>, we are unable to exclude the possibility of a relatively smaller contribution arising from activation of thalamostriatal axons that have distal terminals in dorsolateral striatum nearby to the stimulated region. Thus, although we refer to our measurements as primarily reflecting corticostriatal transmission, our measurements are not 'pure' corticostriatal responses. Borosilicate glass recording electrodes filled with 2 M NaCl were placed in the dorsolateral striatum approximately 400–450  $\mu$ m away from the stimulating electrode. Corticostriatal field population spikes were evoked with 0.15 ms step depolarizations at 0.5 mA intensity at a frequency of 0.05–0.1 Hz. Paired pulses were evoked with a 100 ms inter-stimulus interval. Baseline responses were monitored to ensure stable population spike amplitude for a minimum of 5 min. Input-output functions were then determined for the negative peak 1 (NP1; presynaptic fibre volley) and population spike amplitude by three consecutive rounds of stimulation from 0–1.0 mA in 0.1 mA increments. All recordings were performed at room temperature and acquired using pCLAMP 10 software (Axon Instruments/Molecular Devices). Data analysis was performed blind to genotype in Clamp fit (Axon Instruments/Molecular Devices). Population spike amplitude was measured as the average of the early peak positivity to the peak negativity and from the peak negativity to the late peak positivity. This standard method takes into account the fact that the downward population spike is superimposed on an upward field excitatory postsynaptic potential (fEPSP). Paired pulse ratio (PPR) was calculated as the ratio of the 2nd population spike amplitude to the 1st population spike amplitude for responses to paired pulse stimulation at 0.5 mA fixed intensity with a 100 ms inter-stimulus interval for the pair.

Extracellular field recordings and whole-cell mEPSC recordings in the hippocampal CA1 region were conducted in 300  $\mu$ m thick acute brain slices from 6–9-week-old WT and *Shank3B* mutant mice. For measurement of hippocampal CA1 population spikes, a concentric bipolar stimulating electrode was placed in the striatum radiatum to stimulate the Schaffer collateral pathway, and a borosilicate glass recording electrode ( $\sim 2$ –3 M $\Omega$ ) filled with recording aCSF was placed in the CA1 pyramidal cell layer approximately 400  $\mu$ m from the stimulation site. The recording electrode was placed at the depth in the slice that gave the largest population spike amplitude, and a stable baseline was established for  $< 10$  min. Input-output recordings were conducted by increasing the stimulation intensity from 0 to 160  $\mu$ A in 20  $\mu$ A increments. Three successive rounds were collected and values at each intensity represent the average of the three measurements. CA1 population spike amplitude was quantified exactly as described previously for cortico-striatal population spikes. For CA1 pyramidal neuron whole-cell recordings, pyramidal neurons in CA1 were identified under infrared-differential interference contrast (IR-DIC) visualization. Cells were patched with a Caesium-gluconate-based internal solution containing (in mM): 110 Caesium-gluconate, 15 KCl, 4 NaCl, 5 TEA-Cl, 20 HEPES, 0.2 EGTA, 5 lidocaine *N*-ethyl chloride, 4 ATP magnesium salt, and 0.3 GTP sodium salt. The pH was adjusted to 7.25 with D-gluconic acid and osmolality was adjusted to 290–300 mOsm with sucrose as necessary. The recording aCSF contained 1  $\mu$ M TTX, 100  $\mu$ M picrotoxin, 5  $\mu$ M CGP55845, and 50  $\mu$ M D-APV to isolate pure AMPAR-mediated mEPSCs. CA1 neurons were voltage-clamped at  $-80$  mV to amplify the smallest spontaneous miniature synaptic events that might otherwise escape detection. Criteria for acceptance were uncompensated stable  $R_a < 25$  M $\Omega$  and holding current  $< -300$  pA. mEPSCs were detected using MiniAnalysis software (Synaptosoft) as described for striatal MSNs. All recordings were carried out at room temperature (23–25 °C). Slices were prepared in a 20–30 degree off-horizontal cutting angle (optimal for CA1 region) from one WT and one KO pair each day and the experimenter was blind to the genotypes of the animals.

**Striatal slice preparation for whole-cell recording.** Mice 5–6-week-old were used for all whole-cell electrophysiology procedures by an experimentalist blinded to genotype. Acute coronal striatal slices were prepared as follows. Briefly, mice were anesthetized with Avertin solution (20 mg/ml, 0.5 mg/g body weight) and perfused through the heart with a small volume (about 20 ml) of ice-cold and oxygenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>) cutting solution containing (mM): 105 NMDG, 105 HCl, 2.5 KCl, 1.2 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 25 Glucose, 10 MgSO<sub>4</sub>, 0.5 CaCl<sub>2</sub>, 5 L-Ascorbic Acid, 3 Sodium Pyruvate, 2 Thiourea (pH 7.4, with osmolality of 295–305 mOsm). The brains were rapidly removed and placed in ice-cold and oxygenated cutting solution. The coronal slices (300  $\mu$ m) were prepared



using a slicer (Vibratome 1000 Plus, Leica Microsystems, USA) and then transferred to an incubation chamber (BSK4, Scientific System Design Inc., USA) at 32 °C with carbogenated cutting solution, which was gradually replaced with aCSF in 30 min through a peristaltic pump (Rainin, RP-1) allowing a precise regulation of flowing rates. The slices were then kept in the aCSF that contained (mM): 119 NaCl, 2.3 KCl, 1.0 NaH<sub>2</sub>PO<sub>4</sub>, 26 NaHCO<sub>3</sub>, 11 Glucose, 1.3 MgSO<sub>4</sub>, 2.5 CaCl<sub>2</sub> (pH was adjusted to 7.4 with HCl, with osmolarity of 295–305 mOsm) at room temperature for at least 30 min.

**Whole-cell patch-clamp.** The slice was placed in a recording chamber (RC-27L, Warner Instruments) and constantly perfused with oxygenated aCSF at 24 °C (TC-324B, Warner Instruments) at a rate of 1.5–2.0 ml min<sup>-1</sup>. The striatum and individual MSNs were visualized and identified with a microscope equipped with IR-DIC optics (BX-51WI, Olympus) by location, shape and size (ovoid cell body with major axis of 10 to 14 µm). Two additional measures were used to distinguish them from similar sized GABAergic interneurons. First, GABAergic interneurons show smaller membrane capacitance ( $C_m$ ) and membrane time constant ( $\tau_m$ ) (at least two times less) when compared to that of MSNs. In the case of recordings done with Cs<sup>+</sup> internal, these membrane properties were measured immediately after membrane rupture when the Cs<sup>+</sup> internal has not been dialysed and taken effect yet. Second, AMPA receptor-mediated mEPSCs showed much faster kinetics (including both rise time and decay time constant,  $\tau$  decay) in GABAergic interneurons. Whole-cell patch-clamp recordings were obtained from MSNs using recording pipettes (King Precision Glass, glass type 8250) pulled in a horizontal pipette puller (P-87, Sutter Instruments) to a resistance of 3–4 MΩ, when filled with the internal solution containing (in mM): 107 CsMeSO<sub>3</sub>, 10 CsCl, 3.7 NaCl, 5 TEA-Cl, 20 HEPES, 0.2 EGTA, 5 lidocaine *N*-ethyl chloride, 4 ATP magnesium salt, and 0.3 GTP sodium salt. pH was adjusted to 7.3 with KOH and osmolarity was adjusted to 298–300 mOsm with 15 mM K<sub>2</sub>SO<sub>4</sub>.

To record AMPA receptor-miniature excitatory postsynaptic currents (mEPSCs), the cells were held in voltage clamp at -70 mV in the presence of 50 µM APV (DL-2-amino-5-phosphono-valeric acid), 25 µM BMR (1(S),9(R)-(-)-buciculline methiodide), 10 µM D-serine and 1 µM TTX (all from Tocris). The miniature events were not recorded until 5 min after entering whole cell patch clamp recording mode to allow the dialysis of Cs<sup>+</sup> internal solution for a relatively complete block of the potassium channels in the MSNs. The mEPSCs were detected and analysed with MiniAnalysis (Synaptosoft).

For paired-pulse stimulation experiments, AMPAR mediated excitatory postsynaptic currents (EPSCs) were evoked by a local concentric bipolar stimulating electrode (CBARC75, FHC) that was placed in the inner edge of corpus callosum within the dorso-lateral region of the striatum. Recordings were made in the presence of picrotoxin (100 µM) and APV (50 µM) to block activation of GABA<sub>A</sub> receptors and NMDA receptors. Stimulation was current-controlled (ISO-Flex, A.M.P.I.). The stimulus intensity was set at a level that could evoke 300–400 pA AMPAR-mediated response for all the cells measured and delivered with an inter-stimulus interval of 50 ms. The paired-pulse measurements were obtained for 15–20 consecutive traces and only those traces with stable evoked first current response were used for data analysis. The PPR was calculated with the peak current response to the second pulse divided by that of the first response.

NMDAR- and AMPAR- mediated synaptic current ratio (NMDA/AMPA ratio) was recorded in the presence of picrotoxin at holding potentials of +40 mV and -70 mV, respectively. The NMDA/AMPA ratios were measured according to previously described methods<sup>43</sup>. Briefly, the stimulus intensity was set at a level that could evoke 300–400 pA AMPAR-mediated response with a holding potential at -70 mV. Each evoked response was repeated for 15–20 times with an inter-stimulus interval of 20 s for all the cells measured. The time point of the peak current at -70 mV, considered to be fully mediated by AMPARs, was used to establish the time window for measuring the AMPA peak at +40 mV. The decay to baseline of the AMPA current at -70 mV was used to select a time window for measurement of the NMDA current; a 10-ms measurement window beginning 40 ms after the stimulus artefact was used. This current amplitude at this point was designated as the NMDAR mediated synaptic current response. ( $I_{NMDA}$  at +40 mV/ $I_{AMPA}$  at -70 mV) was taken as the NMDA/AMPA ratio.

**Data acquisition and analysis.** A Multiclamp 700B amplifier (Molecular Devices Corporation) and digidata 1440A were used to acquire whole cell signals. The signals were acquired at 20 kHz and filtered at 2 kHz. The series-resistance was <20 MΩ. Values are expressed as means ± s.e.m. Data were tested for significance using either an unpaired *t*-test or a two-way repeated measures ANOVA.

**Cell filling.** Mice were assigned a code previous to dissection, as to maintain a blinded genotype across all procedures, including dissection, cell filling, imaging and quantification. Mice were deeply anesthetized with an overdose of isoflurane and transcardially perfused with PBS (pH 7.4) followed by ice-cold 4% paraformaldehyde/PBS (PFA) (pH 7.4). The brains were removed and post-fixed

overnight in PFA 4%. After post-fixation, the brain was sliced at 200-µm thickness coronal sections in a vibratome and kept in PBS at 4 °C. For cell filling injections, selected brain slices immersed in PBS were mounted in a tissue stage. Dorsal striatal medium cells were targeted with post hoc confirmation of being medium spiny neurons (morphology and spine density). Using a micromanipulator, micropipettes loaded with Lucifer Yellow dye (Sigma L-0259, 8% solution in 0.05 M Tris buffer, pH 7.4) were used to impale the cell body. A micropipette containing a solution of 0.1 M LiCl was used to deliver the dye with a continuous 10 nA current for 5 min. Following cell filling, a post-staining was used to amplify the fluorescent signal. Briefly, sections were transferred to blocking solution (5% sucrose, 2% BSA, and 1% Triton X-100 in PBS) containing 1:500 rabbit anti-Lucifer Yellow antibody (Invitrogen A5750) and incubated gently for 3 days at 4 °C. Sections were washed three times for 5 min in blocking solution and incubated 2 h at room temperature with 1:400 biotinylated goat anti-rabbit antibody (Vector Laboratories BA-1000). Next, sections were washed three times for 5 min in PBS. A tertiary incubation was performed by incubating sections for 2 h at room temperature in streptavidin-conjugated Alexa 488 (Invitrogen S11223) diluted 1:1,000 in PBS. Finally, sections were washed three times in PBS, mounted on slides using Fluoro-Gel (EMS, 17985-10) and imaged by confocal microscopy. Spine density was calculated automatically using NeuronStudio (Mount Sinai School of Medicine) and manually curated by an observer using a three-dimensional analysis of the dendritic image stack. All spine counts began 30 µm away from the outer edge of the soma and extended for an additional 10–60 µm away from the starting point. The data from spine density passed the Lilliefors normality test and D'Agostino & Pearson omnibus normality test. Spine metrics relating to spine length, spine neck diameter and spine neck width were collected using ImageJ (NIH). All analyses of spine metrics were performed by observers that were blinded to the genotypes of the animals.

**Electron microscopy.** Mice were assigned a code previous to dissection, as to maintain a blinded genotype across all procedures, including dissection, sample processing, imaging and quantification. Mice were deeply anesthetized with an overdose of isoflurane and transcardially perfused with PBS (pH 7.4) followed by ice-cold 4% paraformaldehyde (PFA) in phosphate buffer (pH 7.4). The brains were removed, the striatum dissected and post-fixed overnight in PFA 4%, then transferred into a 4% glutaraldehyde solution and kept at 4 °C for 3 days. The samples were washed twice, 20 min each, in 7.5% sucrose, 0.1 M sodium cacodylate buffer, then post-fixed in 1% osmium tetroxide for 2 h with initial microwave treatment for 6 min. Next, the samples were washed twice in 0.11 M veronal acetate buffer for 20 min each. Following en-block staining in 1% uranyl acetate in distilled water for 1 h the samples were washed twice in 0.11 M veronal acetate buffer for 20 min each. Samples were dehydrated using serial dilutions of ethanol (70%, 95%, 2 × 100%) for 20 min each, with initial microwave treatment of 2 min. Samples were then treated for 20 min twice with propaline oxide and impregnated with 50:50 propaline oxide:Epon resin overnight at 4 °C, with initial microwave treatment for 3 min. Next, the samples were impregnated with 100% Epon resin, three changes of 2 h each, with initial microwave treatment for 3 min each. Tissue samples were embedded in moulds and incubated for 48 h at 60 °C. Afterwards, semi-thin sections (0.5 µm) were cut on a Leica UltraCut S ultramicrotome and stained with Toluidine (0.8%) stain. From these, thin striatal sections (70 nm) were cut on an UltraCut S, mounted on 200 mesh Metaxaform Copper Rhodium grids and post-stained in 2% uranyl acetate in distilled water for 15 min and Sato's Lead citrate stain for 7 min. Grids were examined on a Philips (FEI) CM 12 transmission electron microscope. Images were acquired at ×40,000 magnification using an AMT 2Vue system, with an ORCA HR High resolution digital camera 7 megapixels, a Hamamatsu DCAM board for acquisition and AMT Image Capture Engine software version 600.335f. Images were saved as 7.5 megapixels 8 bit TIFF format files. PSD measurements were performed using ImageJ (NIH) by an observer that was blinded to the genotype of the samples.

**Magnetic resonance image acquisition.** Animals were assigned a blinding code, which was maintained during magnetic resonance (MR) data acquisition and analysis. MR mouse brain imaging was performed on a 7T Bruker Biospec 70/30 horizontal bore system (Billerica). Animals were lightly anesthetized under isoflurane with continuous monitoring and maintenance of physiological parameters throughout the imaging session (~60 min for each animal). Axial two-dimensional T2-weighted fast spin echo images (TURBO-RARE, TE/TR = 11/4,200 ms with 1 mm slice thick, matrix = 256 × 256 and FOV of 2.4 cm × 2.4 cm, five averages, 0.0 mm interslice gap) images were first obtained for screening purposes and supplemental anatomic information. For directed striatal and brain volumetric analysis, 64 contiguous 500-µm thick three-dimensional FSE proton density images (TURBO-RARE, TE/TR = 9/1,500 ms, matrix = 256 × 256 × 64 and FOV 2.2 cm × 2.2 cm × 2.2 cm, 25 min duration) were acquired.

**MR volumetry measurements.** Volumetric analysis of MR data sets was performed in OsiriX software, an open source image processing application

developed and maintained by Pixmeo. The left caudate-putamen was manually segmented in each animal by an investigator blinded to genotype. Each caudate-putamen was traced on contiguous axial slices from the three-dimensional volume acquisition with reference to a high-resolution age-matched mouse brain atlas (The Mouse Brain Atlas, The Mouse Brain Library at <http://www.mbl.org>). Selected areas were reviewed for consistency on coronal and sagittal representations, and cross-correlated with axial two-dimensional FSE images. Volumes were computed within OsiriX. Two separate striatal segmentations were obtained for each animal, with the average volume then taken. Intrarater reliability (kappa value) was = 0.97.

**Volume normalization.** Unilateral caudate-putamen volumes were normalized to brain volume measurements obtained from the same three-dimensional volume sets. Because of susceptibility distortions within the posterior fossa, and variable inclusion from animal to animal of posterior fossa contents at the caudal end of the 64 slice three-dimensional volume set, 'whole' brain volumes for normalization were obtained in each animal from traces beginning rostrally at the

olfactory bulbs and ending caudally through the cerebral aqueduct at the roof of the fourth ventricle. As with striatal volumes, brain volumes were computed in OsiriX from the average of two segmentations. Intrarater reliability (kappa value) was >0.99. Statistical analysis was performed with unpaired two-tailed *t*-test.

39. Welch, J. M., Wang, D. & Feng, G. Differential mRNA expression and protein localization of the SAP90/PSD-95-associated proteins (SAPAPs) in the nervous system of the mouse. *J. Comp. Neurol.* **472**, 24–39 (2004).
40. Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nature Protocols* **1**, 848–858 (2006).
41. Yin, H. H., Davis, M. I., Ronesi, J. A. & Lovinger, D. M. The role of protein synthesis in striatal long-term depression. *J. Neurosci.* **26**, 11811–11820 (2006).
42. Malenka, R. C. & Kocsis, J. D. Presynaptic actions of carbachol and adenosine on corticostriatal synaptic transmission studied *in vitro*. *J. Neurosci.* **8**, 3750–3756 (1988).
43. Myme, C. I., Sugino, K., Turrigiano, G. G. & Nelson, S. B. The NMDA-to-AMPA ratio at synapses onto layer 2/3 pyramidal neurons is conserved across prefrontal and visual cortices. *J. Neurophysiol.* **90**, 771–779 (2003).

# Loss-of-function mutations in sodium channel Na<sub>v</sub>1.7 cause anosmia

Jan Weiss<sup>1\*</sup>, Martina Pyrski<sup>1\*</sup>, Eric Jacobi<sup>1</sup>, Bernd Bufe<sup>1</sup>, Vivienne Willnecker<sup>2</sup>, Bernhard Schick<sup>2</sup>, Philippe Zizzari<sup>3</sup>, Samuel J. Gossage<sup>4</sup>, Charles A. Greer<sup>5</sup>, Trese Leinders-Zufall<sup>1</sup>, C. Geoffrey Woods<sup>6</sup>, John N. Wood<sup>4,7</sup> & Frank Zufall<sup>1</sup>

**Loss of function of the gene *SCN9A*, encoding the voltage-gated sodium channel Na<sub>v</sub>1.7, causes a congenital inability to experience pain in humans. Here we show that Na<sub>v</sub>1.7 is not only necessary for pain sensation but is also an essential requirement for odour perception in both mice and humans. We examined human patients with loss-of-function mutations in *SCN9A* and show that they are unable to sense odours. To establish the essential role of Na<sub>v</sub>1.7 in odour perception, we generated conditional null mice in which Na<sub>v</sub>1.7 was removed from all olfactory sensory neurons. In the absence of Na<sub>v</sub>1.7, these neurons still produce odour-evoked action potentials but fail to initiate synaptic signalling from their axon terminals at the first synapse in the olfactory system. The mutant mice no longer display vital, odour-guided behaviours such as innate odour recognition and avoidance, short-term odour learning, and maternal pup retrieval. Our study creates a mouse model of congenital general anosmia and provides new strategies to explore the genetic basis of the human sense of smell.**

The inability to sense odours is known as general anosmia; individuals born with this phenotype are afflicted with congenital general anosmia. Except for some syndromic cases such as Kallmann syndrome, no causative genes for human congenital general anosmia have been identified so far<sup>1–3</sup>. Nine mammalian genes encoding voltage-gated sodium channel  $\alpha$ -subunits have been cloned and shown to be differentially expressed in the nervous system<sup>4,5</sup>. Of these, *SCN9A*, encoding the tetrodotoxin (TTX)-sensitive sodium channel Na<sub>v</sub>1.7, has received specific attention because of its key role in human pain perception. Individuals carrying loss-of-function mutations in *SCN9A* are unable to experience pain, and an essential requirement of Na<sub>v</sub>1.7 function for nociception in humans has been established<sup>6–9</sup>. Whether all other sensory modalities are fully preserved in these individuals remained unclear, although an association between congenital inability to experience pain and sense of smell deficits has been suggested<sup>7</sup>. In this study we examine human patients carrying *SCN9A* loss-of-function mutations and demonstrate that they fail to sense odours. We establish a mouse model of congenital general anosmia and provide mechanistic insight into the role of Na<sub>v</sub>1.7 in olfaction. Together with previous findings<sup>6–8,10</sup>, our results establish that loss-of-function mutations in a single gene, *SCN9A*, cause a general loss of two major senses—nociception and smell—thus providing a mechanistic link between these two sensory modalities.

## Requirement for Na<sub>v</sub>1.7 in human olfaction

Three individuals with congenital analgesia were ascertained and studied. All three were in their third decade of life and had never experienced acute pain but had no other neurological, cognitive, growth, appearance or health problems. All had suffered from multiple painless fractures and other injuries. Two had given birth painlessly. A working diagnosis of channelopathy-associated insensitivity to pain (CAIP) was made and in each *SCN9A* was sequenced<sup>6</sup>. In the first,

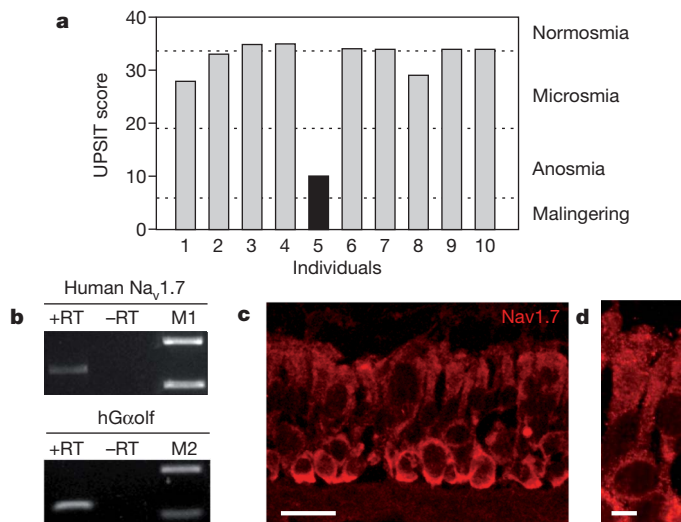
who has been the subject of a detailed case report, the mutations c.774\_775delGT and c.2488C>T were found<sup>10</sup>. These mutations, frameshift and nonsense, respectively, would be predicted to lead to a lack of functional Na<sub>v</sub>1.7 protein. The other two were siblings and had the mutations c.4975A>T and c.3703delATAGCATATGG; again, nonsense and frameshift mutations and predicted to lead to no functional Na<sub>v</sub>1.7 protein. The mother of the siblings was found to be heterozygous for the 11-base-pair deletion and the father heterozygous for the nonsense mutation. Therefore the diagnosis of CAIP was substantiated. We next assessed their sense of smell; none complained of having no sense of smell, one had been a cigarette smoker, none had chronic nasal problems. In the first woman smell function was assessed by using the University of Pennsylvania Smell Identification Test (UPSIT), a standardized 40-item smell test. The results revealed that she was unable to detect any of the odours (Fig. 1a, black bar). Nine healthy, young individuals served as controls (Fig. 1a, grey bars). In the sibling pair we assessed the parents and their two affected offspring together. All were tested in sequence with cotton wool pads suffused with selected odour stimuli: balsamic vinegar, orange, mint, perfume, water (control) and coffee. Both parents correctly identified all stimuli, including smelling nothing for the water. The siblings detected none of the odours. For the siblings the test was repeated using subjectively unpleasant amounts of balsamic vinegar and perfume: the parents identified the odours correctly and found them unpleasant; the siblings neither identified the odours nor experienced any discomfort.

We proposed that these odour-sensing deficits are caused by loss of Na<sub>v</sub>1.7 function in olfactory sensory neurons (OSNs). Indeed, when we investigated expression of Na<sub>v</sub>1.7 in normal human olfactory epithelium, we detected messenger RNA for Na<sub>v</sub>1.7 and the GTP-binding protein G $\alpha_{olf}$ , a prototypical signature of classical OSNs (Fig. 1b). Immunohistochemistry using an antibody specific to Na<sub>v</sub>1.7 verified that Na<sub>v</sub>1.7 is normally expressed in human OSNs (Fig. 1c, d).

<sup>1</sup>Department of Physiology, University of Saarland School of Medicine, 66421 Homburg, Germany. <sup>2</sup>Department of Otolaryngology, University of Saarland School of Medicine, 66421 Homburg, Germany. <sup>3</sup>Centre de Psychiatrie & Neurosciences, UMR 894 Inserm, Faculté de Médecine, Université Paris Descartes, 75014 Paris, France. <sup>4</sup>Molecular Nociception Group, Wolfson Institute for Biomedical Research, University College London, London WC1E 6BT, UK. <sup>5</sup>Department of Neurosurgery, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>6</sup>Department of Medical Genetics, Cambridge Institute for Medical Research, Wellcome/MRC Building, Addenbrooke's Hospital, Cambridge CB2 0XY, UK. <sup>7</sup>Department of Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul 151-742, Korea.

\*These authors contributed equally to this work.





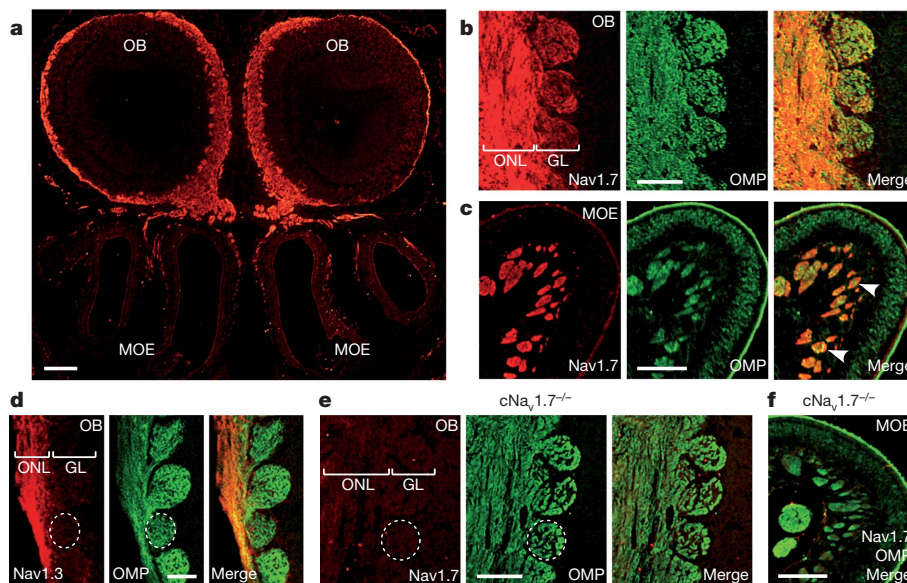
**Figure 1 |  $Na_v1.7$  in human olfaction.** **a**, Quantified olfactory assessment of the first individual with confirmed  $Na_v1.7$  loss-of-function mutations (black bar) using the standardized, 40-item UPSIT test showed that she was unable to detect any of the odour stimuli; the test score revealed general anosmia in this patient. Nine healthy, young individuals served as controls (grey bars). We assessed odour perception in two other individuals with confirmed  $Na_v1.7$  loss-of-function mutations and both were unable to sense any of the odours. These results are described in the main text. **b**, Expression of  $Na_v1.7$  in olfactory epithelium from unaffected normal humans. RT-PCR products with gene-specific primers for human  $Na_v1.7$  (top; size, 1,128 bp) and the G-protein  $G\alpha_{olf}$  (bottom; size, 143 bp). PCRs were carried out with equal amounts of RNA in the presence (+RT) or absence (–RT) of reverse transcriptase to exclude product amplification from genomic DNA. M1, size marker; 2,000 bp, 850 bp; M2, size marker; 400 bp, 100 bp. Similar results were obtained in two other human olfactory mucosa samples. **c**, Confocal fluorescence image of  $Na_v1.7$  immunoreactivity (red) in a cryosection of human olfactory epithelium. Scale bar, 20  $\mu$ m. **d**, Enlargement showing a single OSN. Scale bar, 5  $\mu$ m.

## Conditional $Na_v1.7$ null mice

To investigate the mechanisms that underlie the essential role of  $Na_v1.7$  in odour perception, we first examined  $Na_v1.7$  expression in the mouse olfactory system and then used the Cre-loxP system to delete the channel in those cells that express olfactory marker protein (OMP), which includes all classical OSNs<sup>11</sup>. These mice enabled us to examine the mechanisms underlying  $Na_v1.7$ -associated anosmia and the behavioural consequences. Consistent with our findings in human OSNs, OSNs from wild-type mice (C57BL/6, referred to as B6) showed  $Na_v1.7$  immunoreactivity at their somata (Supplementary Fig. 1). Of greater interest, coronal sections containing main olfactory epithelium (MOE), olfactory nerves and the two olfactory bulbs revealed the most marked  $Na_v1.7$  staining in the superficial olfactory nerve layer (ONL, containing axons from OSNs) as well as the glomerular layer (a complex neuropil that includes the presynaptic OSN boutons) of the olfactory bulb (Fig. 2a–c). Higher magnification of individual glomeruli verified co-expression of  $Na_v1.7$  with OMP in the glomerular neuropil (Fig. 2b), whereas olfactory bulb projection neurons (the mitral/tufted or M/T cells) and local interneurons did not show  $Na_v1.7$  immunoreactivity (Fig. 2a). Thus,  $Na_v1.7$  occupies a critical presynaptic location at the first synapse in the olfactory system.

$Na_v1.7$  is not the sole  $Na_v$  channel expressed in mouse OSNs. Real-time quantitative polymerase chain reaction with reverse transcription (qRT-PCR) analysis identified  $Na_v1.3$  as an additional candidate (Supplementary Fig. 2) and immunohistochemistry verified its expression in OSNs and their axons (Fig. 2d). However, unlike  $Na_v1.7$  we did not observe  $Na_v1.3$  immunoreactivity in individual glomeruli (Fig. 2d), indicating that  $Na_v1.7$  could be the sole  $Na_v$  channel underlying action potential propagation in olfactory glomeruli and OSN nerve terminals.

To create a conditional knockout mouse model, we crossed ‘floxed’  $Na_v1.7$  mice harbouring a loxP-flanked *Scn9a* gene<sup>12</sup> to homozygous OMP-Cre mice in which the OMP-coding region is replaced by that of Cre recombinase<sup>13</sup>. Further breeding established offspring that were both homozygous for the floxed *Scn9a* alleles and heterozygous for *cre* and *Omp*. In these mice, Cre-mediated  $Na_v1.7$  deletion was restricted to OMP-positive cells (henceforward referred to as  $cNa_v1.7^{-/-}$  mice). These mice lacked  $Na_v1.7$  expression in a tissue-specific manner



**Figure 2 |  $Na_v1.7$  expression in the mouse main olfactory system.** **a**,  $Na_v1.7$  immunoreactivity (red) detected in the olfactory bulbs (OB, top) and the underlying main olfactory epithelium (MOE, bottom). B6 mouse, 2-days old. **b**, Colocalization of  $Na_v1.7$  (red) and OMP (green) expression in the ONL and glomerular layer (GL) of the olfactory bulb. B6 mouse, 3-weeks old. **c**, Strong  $Na_v1.7$  immunoreactivity in OSN axon bundles (arrowheads). B6 mouse,

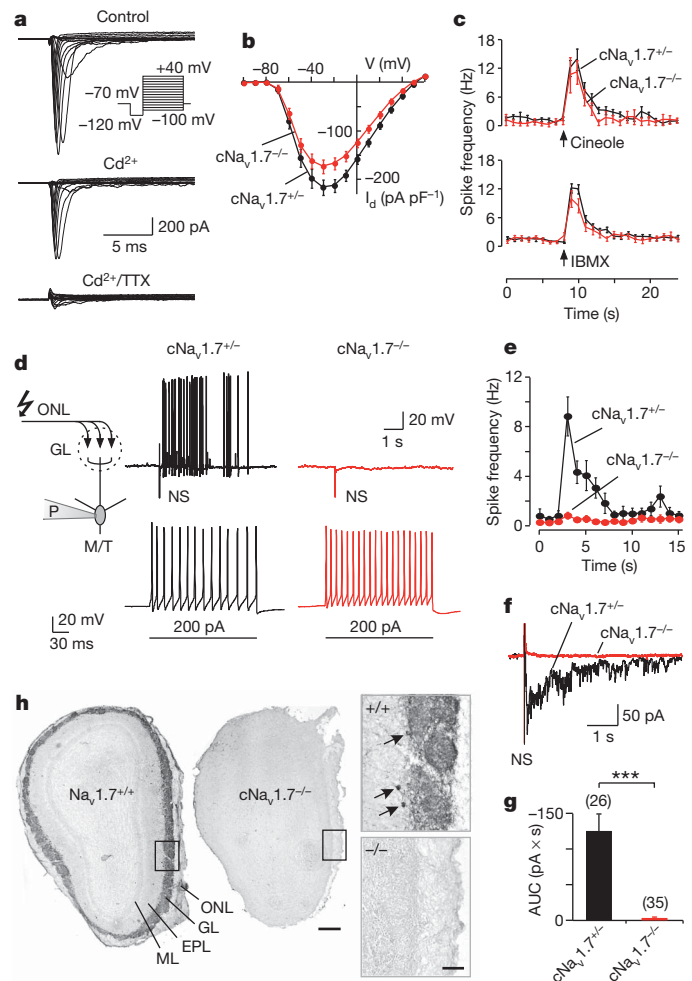
3-weeks old. **d**,  $Na_v1.3$  expression (red) terminates in the ONL and is not detectable in OMP-labelled (green) glomeruli. OMP-GFP mouse, 16-days old. **e–f**,  $cNa_v1.7^{-/-}$  mice lack  $Na_v1.7$  immunoreactivity in the ONL and glomerular layer of the olfactory bulb (**e**) as well as in OSN axon bundles of MOE (**f**). Dashed circle, individual glomerulus. Scale bars: **a**, 200  $\mu$ m; **b**, **c**, **e**, **f**, 100  $\mu$ m; **d**, 50  $\mu$ m.

(Fig. 2e, f and Supplementary Fig. 3). Successful matings occurred between  $cNa_v1.7^{+/-}$  males and  $cNa_v1.7^{-/-}$  females whereas homozygous knockout pairs did not produce any offspring.  $cNa_v1.7^{-/-}$  mice showed a reduced body weight during the first three months of postnatal development (Supplementary Fig. 4a). Because both  $Na_v1.7$  (refs 14, 15) and  $OMP^{16}$  are also expressed in some neurons mediating hormonal regulation, we assayed insulin-like growth factor (IGF-1, also known as somatomedin C) in  $cNa_v1.7^{-/-}$  ( $650 \pm 94$  ng  $ml^{-1}$ ;  $n = 4$ ) versus  $cNa_v1.7^{+/-}$  mice ( $684 \pm 27$  ng  $ml^{-1}$ ;  $n = 4$ ; mean  $\pm$  s.d.) but found no significant difference between the two genotypes ( $P = 0.26$ ). Given that newborn  $cNa_v1.7^{-/-}$  mice had very little milk in their stomachs (Supplementary Fig. 4b), the diminished weight gain was probably caused by a deficit to suckle effectively, consistent with results in mice deficient in  $G\alpha_{olf}$  (*Gnal*) or the cAMP-gated cation channel (*Cnga2*)<sup>17,18</sup>.

### Loss of synaptic transfer in olfactory glomeruli

To define the function of  $Na_v1.7$  in OSNs, we prepared MOE tissue slices<sup>19</sup> and recorded sodium currents in voltage-clamped OSNs. Both  $Na_v1.7^{+/-}$  and  $Na_v1.7^{-/-}$  OSNs displayed sizeable, TTX-sensitive sodium currents in response to step depolarizations (Fig. 3a, b). On the basis of its biophysical properties,  $Na_v1.7$  has been suggested to transduce generator potentials into action potentials in sensory neurons<sup>9</sup>. However, peak current densities of voltage-activated sodium currents were reduced only moderately, by about 20%, in  $Na_v1.7^{-/-}$  OSNs (Fig. 3b). To determine whether  $Na_v1.7^{-/-}$  OSNs could still produce odour-evoked action potentials, we used extracellular loose-patch recording from visually identified OSN dendritic knobs<sup>20</sup> and analysed spike frequency histograms after brief odour exposure (Fig. 3c). There was no obvious difference in odour responsiveness in  $Na_v1.7^{+/-}$  versus  $Na_v1.7^{-/-}$  OSNs (Fig. 3c). We obtained similar results when we stimulated the cells with 3-isobutyl-1-methylxanthine (IBMX)<sup>21</sup>, which raises intracellular cAMP by inhibiting endogenous phosphodiesterase activity (Fig. 3c). Thus, although the initial site of odour-evoked action potential generation in OSNs is unknown,  $Na_v1.7$  is not essential for this activity.

Because  $Na_v1.7$  is expressed in olfactory bulb glomeruli (Fig. 2), we reasoned that it could be required for action potential conduction in OSN terminals. Olfactory glomeruli are delineated spheres of neuropil containing synapses from the OSN axon terminals onto juxtaglomerular interneurons and M/T projection neurons<sup>22,23</sup>. To examine whether presynaptic activity of  $Na_v1.7$  underlies transmitter release in the olfactory glomerulus, we prepared olfactory bulb tissue slices<sup>24</sup> and combined ONL focal electric stimulation with whole-cell patch-clamp recording from visually identified M/T cells. With the chosen protocol, in control  $cNa_v1.7^{+/-}$  mice a single electrical stimulus in the ONL produced a reliable postsynaptic response in M/T cells. Under current clamp, such responses consisted of a prolonged excitation lasting on average for  $2.4 \pm 0.4$  s (Fig. 3d, top;  $n = 29$ ), with response latencies of  $22 \pm 4$  ms ( $n = 29$ ). Under voltage clamp, we observed bursts of postsynaptic currents (Fig. 3f; duration,  $3.2 \pm 0.4$  s;  $n = 26$ ). In stark contrast, in the  $cNa_v1.7^{-/-}$  mice such postsynaptic responses were virtually absent in M/T cells, even when the stimulus strength was increased by several-fold (Fig. 3d–f;  $n = 49$ ). Importantly, M/T cells in these mice still produced normal action potentials when depolarized via current injection through the patch pipette (Fig. 3d, bottom), consistent with the fact that M/T cells lack both  $OMP$  and  $Na_v1.7$  expression (Fig. 2) and indicating that the effect of deleting  $Na_v1.7$  is presynaptic to the M/T cells. The inability of M/T cells to produce synaptic responses to ONL stimulation was not due to a potential deficit in synapse formation because: (1) immunohistochemistry showed normal expression of the vesicular glutamate transporter 2 (vGluT2, which is selectively expressed in OSN axon terminals)<sup>25,26</sup> (Supplementary Fig. 5); and (2) electron microscopy revealed the existence of normal OSN boutons and synapses in the glomeruli of  $cNa_v1.7^{-/-}$  mice (Supplementary Fig. 6). Furthermore, conditional



**Figure 3 |  $Na_v1.7$  is essential for synaptic transfer in the olfactory glomerulus.** **a**, Families of whole-cell currents to a series of depolarizing voltage steps recorded from  $cNa_v1.7^{-/-}$  OSNs. OSNs were exposed successively to extracellular bath solution (control), bath solution containing  $Cd^{2+}$  (100  $\mu M$ ), and bath solution containing  $Cd^{2+}$  (100  $\mu M$ ) and TTX (2  $\mu M$ ). Holding potential,  $-70$  mV. **b**, Current-density-voltage curves of sodium currents from  $cNa_v1.7^{+/-}$  (black;  $n = 35$ ) and  $cNa_v1.7^{-/-}$  OSNs (red;  $n = 46$ ). Current densities of  $Na_v1.7^{-/-}$  OSNs were significantly diminished between  $-30$  mV and  $10$  mV (LSD,  $P = 0.001$ – $0.04$ ). **c**, Action potential responses in visually identified OSN dendritic knobs to pulsed stimulation (0.2 s) with cineole (100  $\mu M$ ) or IBMX (100  $\mu M$ ) in  $cNa_v1.7^{+/-}$  (black;  $n = 28$  and  $78$  cells, respectively) and  $cNa_v1.7^{-/-}$  OSNs (black;  $n = 5$  and  $25$ , respectively). Firing properties were similar in both genotypes (LSD,  $P = 0.14$ – $0.73$ ). **d**, M/T cells exhibit postsynaptic potentials to presynaptic nerve stimulation (NS) in  $cNa_v1.7^{+/-}$  (black), but not in  $cNa_v1.7^{-/-}$  mice (red), whereas M/T cells in both genotypes show normal action potentials to current injection (200 pA). Current-clamp whole-cell recording. **e**, Absence of M/T cell postsynaptic potentials to nerve stimulation in  $cNa_v1.7^{-/-}$  (red;  $n = 44$ ) versus  $cNa_v1.7^{+/-}$  (black;  $n = 29$ ) (LSD,  $P < 0.01$ – $0.001$ ). **f**, Example of postsynaptic currents in M/T cells of  $cNa_v1.7^{+/-}$  (black) and  $cNa_v1.7^{-/-}$  mice (red). **g**, Analysis of area under curve (AUC) of M/T cell postsynaptic currents during a 5 s interval after nerve stimulation in  $cNa_v1.7^{+/-}$  (black) and  $cNa_v1.7^{-/-}$  mice (red). Number of cells tested is shown in brackets above each bar. Unpaired *t*-test. \*\*\* $P < 0.0001$ . **h**, Marked reduction of TH expression in juxtaglomerular cells of the olfactory bulb in  $cNa_v1.7^{-/-}$  mice. EPL, external plexiform layer; ML, mitral cell layer. Arrows in the inset (+/+) indicate individual juxtaglomerular cells. Scale bars: overview, 100  $\mu m$ ; inset, 50  $\mu m$ . Error bars represent mean  $\pm$  s.e.m.

OSN expression of tetanus toxin light chain, which inhibits synaptic release, does not alter the pattern of axonal targeting in olfactory bulb glomeruli during development<sup>27</sup>.

Tyrosine hydroxylase (TH) expression in juxtaglomerular neurons of the olfactory bulb, a correlate of afferent trans-synaptic activity,



requires olfactory nerve input and odour-stimulated glutamate release by OSN terminals<sup>28</sup>. Consistent with a loss of OSN synaptic release, TH expression was markedly reduced in *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice (Fig. 3h;  $n = 6$ ). The level of TH downregulation was similar to that observed after odour deprivation by naris occlusion<sup>29</sup> or after deletion of the *Cnga2* cation channel gene<sup>30</sup>. Thus, we conclude that the presence of *Na<sub>v</sub>1.7* in OSN axons is an essential and non-redundant requirement to initiate information transfer from OSN terminals to neurons in the olfactory bulb.

### The absence of odour-guided behaviours

To further validate these results, we investigated several odour-guided behaviours in B6, *cNa<sub>v</sub>1.7<sup>+/-</sup>* and *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice. First, we performed an odour preference test<sup>31</sup> to assess recognition abilities for innate odour qualities (Fig. 4a). Filter papers scented with various cues representing both species-specific and food odours (male and female urine, peanut butter, milk) were presented to the mice and investigation times were analysed. Water was used as a neutral stimulus and 1,8-cineole (eucalyptol), which does not evoke innate attraction, served as the control ( $n \geq 7$  for each cue and strain, respectively). B6 and *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice both showed strong attraction towards conspecific and food odours, whereas *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice failed to show any interest in these stimuli.

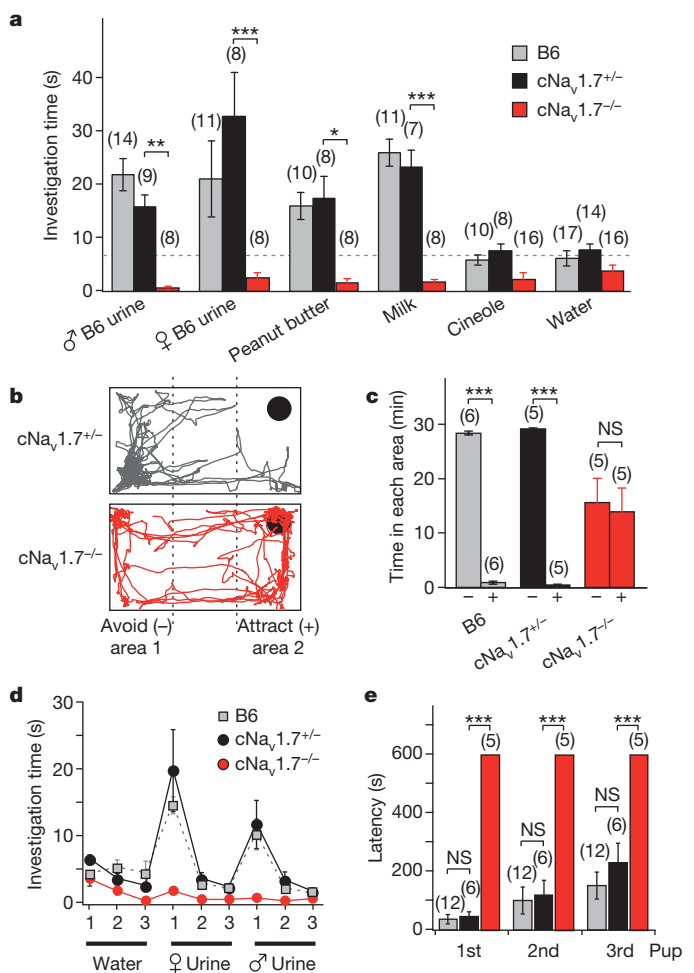
Second, we explored whether *Na<sub>v</sub>1.7* is required for innate avoidance behaviour towards a predator odour, trimethyl-thiazoline (TMT)<sup>31</sup>, which is normally secreted from the fox anal gland and known to induce aversive behaviour and fear responses in mice. We observed robust avoidance behaviour in both B6 ( $n = 6$ ) and *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice ( $n = 5$ ) but, notably, *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice lacked an innately aversive response in this assay ( $n = 5$ ; Fig. 4b, c).

Third, we investigated the performance of *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice in a habituation–dishabituation assay, which allows for measurement of novel odour investigation, short-term odour learning, and odour discrimination<sup>32</sup> (Fig. 4d). Mice of both sexes were each presented three distinct stimuli (water, female urine, male urine), each delivered for three successive trials, and investigation time during each trial (3 min) was analysed. Consistent with the results of Fig. 4a, *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice ( $n = 8$ ) failed to show significant odour investigation, habituation, or discrimination abilities when compared with B6 ( $n = 8$ ) or *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice ( $n = 8$ ) (Fig. 4d; least significant difference (LSD),  $P < 0.0001$ ).

Last, we examined pup retrieval ability of female mice, a social behaviour that probably depends on a functional main olfactory system (Fig. 4e). Three pups of a litter were removed from the nest, randomly distributed in the cage, and the time to retrieve each pup into the nest was quantified. In contrast to the performance of B6 ( $n = 12$ ) or *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice ( $n = 6$ ), *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice ( $n = 5$ ) failed to retrieve any of the three pups during a 10-min trial period (Fig. 4e).

### Conclusions and prospects

Our results establish a critical role of the *Na<sub>v</sub>1.7* sodium channel in olfaction. Using conditional *Na<sub>v</sub>1.7* null mice, we demonstrate that, in the absence of *Na<sub>v</sub>1.7*, OSNs are still electrically active and generate odour-evoked action potentials but fail to initiate synaptic signalling to the projection neurons in the olfactory bulb. These results provide evidence that *Na<sub>v</sub>1.7* is an essential and non-redundant requirement for action potential propagation in the sections of OSN axons within the olfactory glomerulus. The conditional null mice no longer show a wide range of vital, odour-guided behaviours including innate attraction to food and conspecific odours, odour discrimination and short-term odour learning, innate avoidance towards a predator odour, effective suckling behaviour of newborn pups, and maternal pup retrieval. Within the limits of our anatomical analyses, synapse formation in these mice appears normal, indicating that the behavioural phenotype of the mutant mice is most likely the result of a loss of signalling at the first synapse in the olfactory system. Whether *Na<sub>v</sub>1.7* or other sodium channel subunits such as *Na<sub>v</sub>1.3* are involved in OSN



**Figure 4** | *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice are anosmic. **a**, Innate olfactory preference test. Mean investigation times were quantified in B6, *cNa<sub>v</sub>1.7<sup>+/-</sup>* and *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice during a 3-min test period. Mean investigation time for water in B6 and *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice (dashed line) served as a marker for attraction. Number of animals tested is shown in brackets above each bar. LSD: \* $P < 0.02$ ; \*\* $P < 0.001$ ; \*\*\* $P < 0.0001$ . No difference was observed between B6 and *cNa<sub>v</sub>1.7<sup>+/-</sup>* mice (LSD,  $P = 0.1$ ). **b**, **c**, Innate olfactory avoidance to the predator odour TMT. **b**, Examples of the trajectory plot of the position of a *cNa<sub>v</sub>1.7<sup>+/-</sup>* and a *cNa<sub>v</sub>1.7<sup>-/-</sup>* mouse (30-min video-tracking). Location of TMT (5 µl) indicated by the circle in the upper right corner. **c**, Quantification of time spent in either area 1 (–) or area 2 (+) during exposure to TMT indicates that *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice lack avoidance behaviour to TMT (LSD,  $P = 0.67$ ). LSD: \*\*\* $P < 0.0001$ . **d**, Olfactory habituation–dishabituation assay. Mean investigation time of B6, *cNa<sub>v</sub>1.7<sup>+/-</sup>* and *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice ( $n = 8$  for each group) to three distinct stimuli (water, female and male mouse urine) during a 3-min test period were quantified. Numbers indicate stimulus presentation order. **e**, Pup retrieval test. Mean latencies of B6, *cNa<sub>v</sub>1.7<sup>+/-</sup>* and *cNa<sub>v</sub>1.7<sup>-/-</sup>* female mice in retrieving three individual pups that were randomly distributed throughout the cage. *cNa<sub>v</sub>1.7<sup>-/-</sup>* mice failed to retrieve any of the pups during a 10-min trial period (LSD,  $P < 0.0001$ ). NS, not significant (LSD,  $P = 0.41$ ). Error bars represent mean  $\pm$  s.e.m.

axon pathfinding and activity-dependent neural map formation<sup>33</sup> in the mouse olfactory system remains to be seen. Importantly, the phenotype of the mutant mice—the inability to perceive odours—is similar to that observed in human patients with confirmed *Na<sub>v</sub>1.7* loss-of-function mutations. Smell tests in three individuals with congenital analgesia establish that they are unable to sense any of the odours. Systematic olfactory testing of patients carrying *Na<sub>v</sub>1.7* loss-of-function mutations will be required in the future.

The genetic basis of sensory deficits such as blindness, deafness and pain disorders has been extensively studied in recent years. By comparison, relatively little progress has been made in understanding



human congenital general anosmia<sup>1</sup>. Mutations in olfactory signal transduction genes such as *CNGA2*, *GNAL* and *ADCY3* do not seem to be a major cause of human congenital general anosmia<sup>2</sup>. The identification of a sodium channel subunit as a causative gene for an inherited form of general anosmia provides new insight into the molecular pathophysiology of olfaction and should stimulate further research aimed at understanding the genetic basis of the human sense of smell.

## METHODS SUMMARY

**Human subjects.** All research involving humans was obtained with the informed consent of the patients and performed under protocols approved by the Ethics Committee of the relevant institution. Human nasal mucosa was obtained by biopsy during routine nasal surgery. Further details of the human smell tests can be found in Methods.

**Animals.** The relevant Institutional Animal Care and Use Committee approved all procedures. Experiments used tissue-specific Na<sub>v</sub>1.7-deficient, C57BL/6J and OMP-GFP mice. See Methods for details.

**PCR analyses.** Human olfactory mucosa samples were examined individually whereas mouse tissue was pooled from four different animals. PCR products were amplified with gene-specific primers and specificity was controlled by sequencing. Primers are specified in Methods.

**Immunohistochemistry and electron microscopy.** These followed previously published procedures<sup>26,34,35</sup> as described in Methods.

**Electrophysiology.** We used intact mouse MOE preparations as described previously<sup>19,20</sup>. Details of the electrophysiological experiments<sup>36</sup> are given in Methods.

**Behavioural analyses.** Innate olfactory preference tests and avoidance measurements used methods similar to those previously described<sup>31,37</sup>. Olfactory habituation–dishabituation followed previously described procedures<sup>32</sup>. Details are given in Methods.

**IGF-1 assays.** These were done as described in Methods.

**Statistics.** Details of the statistical tests are given in Methods. Unless otherwise stated, results are presented as means ± s.e.m.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 15 November 2010; accepted 4 March 2011.**

**Published online 23 March 2011.**

- Hasin-Brumshtein, Y., Lancet, D. & Olender, T. Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet.* **25**, 178–184 (2009).
- Feldmesser, E. *et al.* Mutations in olfactory signal transduction genes are not a major cause of human congenital general anosmia. *Chem. Senses* **32**, 21–30 (2007).
- Keller, A. & Vosshall, L. B. Better smelling through genetics: mammalian odor perception. *Curr. Opin. Neurobiol.* **18**, 364–369 (2008).
- Goldin, A. L. Resurgence of sodium channel research. *Annu. Rev. Physiol.* **63**, 871–894 (2001).
- Catterall, W. A., Goldin, A. L. & Waxman, S. G. International Union of Pharmacology. XLVII. Nomenclature and structure–function relationships of voltage-gated sodium channels. *Pharmacol. Rev.* **57**, 397–409 (2005).
- Cox, J. J. *et al.* An SCN9A channelopathy causes congenital inability to experience pain. *Nature* **444**, 894–898 (2006).
- Goldberg, Y. P. *et al.* Loss-of-function mutations in the Na<sub>v</sub>1.7 gene underlie congenital indifference to pain in multiple human populations. *Clin. Genet.* **71**, 311–319 (2007).
- Ahmad, S. *et al.* A stop codon mutation in SCN9A causes lack of pain sensation. *Hum. Mol. Genet.* **16**, 2114–2121 (2007).
- Dib-Hajj, S. D., Cummins, T. R., Black, J. A. & Waxman, S. G. From genes to pain: Na<sub>v</sub>1.7 and human pain disorders. *Trends Neurosci.* **30**, 555–563 (2007).
- Nilsen, K. B. *et al.* Corrigendum to “Two novel SCN9A mutations causing insensitivity to pain.” *Pain* **145**, 264 (2009).
- Munger, S. D., Leinders-Zufall, T. & Zufall, F. Subsystem organization of the mammalian sense of smell. *Annu. Rev. Physiol.* **71**, 115–140 (2009).
- Nassar, M. A. *et al.* Nociceptor-specific gene deletion reveals a major role for Na<sub>v</sub>1.7 (PN1) in acute and inflammatory pain. *Proc. Natl Acad. Sci. USA* **101**, 12706–12711 (2004).
- Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428**, 393–399 (2004).
- Klugbauer, N., Lacinova, L., Flockerzi, V. & Hofmann, F. Structure and functional expression of a new member of the tetrodotoxin-sensitive voltage-activated sodium channel family from human neuroendocrine cells. *EMBO J.* **14**, 1084–1090 (1995).
- Morinville, A. *et al.* Distribution of the voltage-gated sodium channel Na<sub>v</sub>1.7 in the rat: expression in the autonomic and endocrine systems. *J. Comp. Neurol.* **504**, 680–689 (2007).
- Baker, H., Grillo, M. & Margolis, F. L. Biochemical and immunocytochemical characterization of olfactory marker protein in the rodent central nervous system. *J. Comp. Neurol.* **285**, 246–261 (1989).
- Belluscio, L., Gold, G. H., Nemes, A. & Axel, R. Mice deficient in G<sub>o</sub><sub>ir</sub> are anosmic. *Neuron* **20**, 69–81 (1998).
- Zhao, H. & Reed, R. R. X inactivation of the *OCNC1* channel gene reveals a role for activity-dependent competition in the olfactory system. *Cell* **104**, 651–660 (2001).
- Spehr, M. *et al.* Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. *J. Neurosci.* **26**, 1961–1970 (2006).
- Leinders-Zufall, T. *et al.* Contribution of the receptor guanylyl cyclase GC-D to chemosensory function in the olfactory epithelium. *Proc. Natl Acad. Sci. USA* **104**, 14507–14512 (2007).
- Munger, S. D. *et al.* Central role of the CNGA4 channel subunit in Ca<sup>2+</sup>-calmodulin-dependent odor adaptation. *Science* **294**, 2172–2175 (2001).
- Shepherd, G. M., Chen, W. R. & Greer, C. A. in *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 165–216 (Oxford Univ. Press, 2004).
- Wachowiak, M. & Shipley, M. T. Coding and synaptic processing of sensory information in the glomerular layer of the olfactory bulb. *Semin. Cell Dev. Biol.* **17**, 411–423 (2006).
- Nickell, W. T., Shipley, M. T. & Behbehani, M. M. Orthodromic synaptic activation of rat olfactory bulb mitral cells in isolated slices. *Brain Res. Bull.* **39**, 57–62 (1996).
- Gabellec, M. M., Panzanelli, P., Sassoe-Pognetto, M. & Lledo, P. M. Synapse-specific localization of vesicular glutamate transporters in the rat olfactory bulb. *Eur. J. Neurosci.* **25**, 1373–1383 (2007).
- Richard, M. B., Taylor, S. R. & Greer, C. A. Age-induced disruption of selective olfactory bulb synaptic circuits. *Proc. Natl Acad. Sci. USA* **107**, 15613–15618 (2010).
- Yu, C. R. *et al.* Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. *Neuron* **42**, 553–566 (2004).
- Puche, A. C. & Shipley, M. T. Odor-induced, activity-dependent transneuronal gene induction *in vitro*: mediation by NMDA receptors. *J. Neurosci.* **19**, 1359–1370 (1999).
- Cho, J. Y., Min, N., Franzen, L. & Baker, H. Rapid down-regulation of tyrosine hydroxylase expression in the olfactory bulb of naris-occluded adult rats. *J. Comp. Neurol.* **369**, 264–276 (1996).
- Baker, H. *et al.* Targeted deletion of a cyclic nucleotide-gated channel subunit (OCNC1): biochemical and morphological consequences in adult mice. *J. Neurosci.* **19**, 9313–9321 (1999).
- Kobayakawa, K. *et al.* Innate versus learned odour processing in the mouse olfactory bulb. *Nature* **450**, 503–508 (2007).
- Wesson, D. W., Levy, E., Nixon, R. A. & Wilson, D. A. Olfactory dysfunction correlates with amyloid- $\beta$  burden in an Alzheimer's disease mouse model. *J. Neurosci.* **30**, 505–514 (2010).
- Sakano, H. Neural map formation in the mouse olfactory system. *Neuron* **67**, 530–542 (2010).
- Pyrski, M. *et al.* Sodium/calcium exchanger expression in the mouse and rat olfactory systems. *J. Comp. Neurol.* **501**, 944–958 (2007).
- Au, W. W., Treloar, H. B. & Greer, C. A. Sublaminar organization of the mouse olfactory bulb nerve layer. *J. Comp. Neurol.* **446**, 68–80 (2002).
- Ukhanov, K., Leinders-Zufall, T. & Zufall, F. Patch-clamp analysis of gene-targeted vomeronasal neurons expressing a defined V1r or V2r receptor: ionic mechanisms underlying persistent firing. *J. Neurophysiol.* **98**, 2357–2369 (2007).
- Papes, F., Logan, D. W. & Stowers, L. The vomeronasal organ mediates interspecies defensive behaviors through detection of protein pheromone homologs. *Cell* **141**, 692–703 (2010).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank the individuals who participated in this study, P. Mombaerts for supplying OMP-Cre and OMP-GFP mice, F. Margolis for anti-OMP antibodies, J. Epelbaum for supporting the IGF-1 measurements, P. Hammes for assistance with the immunohistochemistry and C. Kaliszewski for assistance with the electron microscopy. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) to F.Z. (SFB 530 and SFB 894) and T.L.-Z. (SFB 894). E.J. was supported by the DFG-funded International Graduate School GK 1326. T.L.-Z. is a Lichtenberg Professor of the Volkswagen Foundation. J.N.W. was supported by the Biotechnology and Biological Sciences Research Council, Medical Research Council, Wellcome Trust and grant number R31-2008-000-10103-0 from the World Class University project of the Korean Ministry of Education, Science and Technology and the National Research Foundation of Korea.

**Author Contributions** J.N.W. and F.Z. conceived the study. J.W., M.P., E.J., B.B. and P.Z. performed experiments. B.S. performed human biopsies. C.G.W., S.J.G. and J.N.W. performed human smell tests. J.W., M.P., E.J., B.B., V.W., P.Z., S.J.G., C.A.G., T.L.-Z., C.G.W., J.N.W. and F.Z. analysed results. M.P., T.L.-Z., J.N.W. and F.Z. contributed key reagents. F.Z. wrote the manuscript. All authors edited the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [www.nature.com/nature](http://www.nature.com/nature). Correspondence and requests for materials should be addressed to F.Z. ([frank.zufall@uks.eu](mailto:frank.zufall@uks.eu)).

## METHODS

**Human biopsies.** Human nasal mucosa was obtained by biopsy during routine nasal surgery with patients under general anaesthesia. Biopsy specimens were obtained from three individuals and snap-frozen in liquid nitrogen for later processing. All samples were obtained under a protocol approved by the Ethics Committee of the University of Saarland School of Medicine. All biopsy tissues were obtained with the informed consent of the patients.

**Human psychophysics.** The UPSIT was obtained from Sensonics. The test was applied over a period of 25 min. Testing and scoring was done according to standardized operating procedures summarized in the test manual. The reference values have been derived from recorded reference ranges for the UPSIT test based on British individuals.

**Olfactory mucosa biopsies and PCR analyses.** Human surgical material containing olfactory mucosa collected from three different patients was examined individually. RT-PCRs from human samples were performed on a MyCycler (BIO-RAD) with Herculase (Agilent Technologies) following suppliers' instructions. To amplify human  $G_{\alpha_{olf}}$  we used the oligonucleotides TGGAAAGA ATCGACAGCGTCAGC and GGCCACCAACATCAAAACATGTGG. Human  $Na_v1.7$  was amplified by CATGAATAACCCACCGGACTG and CCTATGCC CTTTCGACACCAAGG. PCR conditions were: 95 °C for 2 min pre-denaturation, followed by 35 cycles (95 °C for 30 s, 60 °C for 30 s ( $G_{\alpha_{olf}}$ ) or 1 min ( $Na_v1.7$ ), 72 °C for 30 s), followed by a final extension 72 °C for 5 min. Mouse tissue was pooled from four different B6 mice (4–8 weeks old). RNA was isolated with the InnuPREP RNA isolation kit (Analyticjena). RNA quality was assessed by gel electrophoresis and photometric measurements. cDNA was synthesized from 0.5 µg of total RNA using the Smart cDNA Synthesis technology (Clontech) and Superscript II reverse transcriptase (Invitrogen). qPCR for different mouse  $Na_v$  subunits were done on a My-iQ-cycler using iQ<sup>TM</sup> SYBRGreen Supermix following the supplier's instructions (BIO-RAD). We used the following oligonucleotides:  $Na_v1.1$  (AGCCTGGTAGAAGCTTGGCCTTGC and TGCCAAACCA CGGCAAAAATAAAG);  $Na_v1.2$  (TGGGATCTTCACCGCAGAAATG and TGGGCCAGGATTTTGCCAAAC);  $Na_v1.3$  (AGCTTGGCCTGGCAAACGTG and ATGCCGACCAACCGCAAAAATG);  $Na_v1.5$  (ACAGCCGAGTTTGAG GAGATGC and CGCTGATTCCGGTGCTCTCA);  $Na_v1.6$  (ACGCCACAATTC GAACATGTCC and CCTGGCTGATCTTACAGACGCA);  $Na_v1.7$  (ACGGAT GAATTCAAAATGTACTTGCAG and GTTCTCGTTGATCTTGCAAACA CA). PCR conditions were: 95 °C for 3 min pre-denaturation, followed by 42 cycles 95 °C for 30 s, 64 °C for 20 s, 72 °C for 30 s. Each reaction was performed in three replicates on 96-well plates and analysed with the iQ5 Software (BIO-RAD). Specificity of all PCR products was confirmed by gel electrophoresis and sequencing.

**Mice.** Animal care and experimental procedures were performed in accordance with the guidelines established by the animal welfare committee of the University of Saarland School of Medicine. Mice were kept under a standard light/dark cycle with food and water *ad libitum*. Tissue-specific,  $Na_v1.7$ -deficient mice were generated by crossbreeding 'floxed'  $Na_v1.7$  mice that carry two *loxP* sites, flanking exons 14 and 15 of *Scn9a*<sup>12</sup> with homozygous OMP-Cre mice (B6;129P2-*Omp*<sup>tm4(Cre)Mom</sup>/MOM) that express Cre recombinase under the control of the OMP promoter<sup>13</sup>. Further breeding established offspring that were both homozygous for the floxed  $Na_v1.7$  alleles and heterozygous for *cre* and *Omp*. In these mice, Cre-mediated  $Na_v1.7$  deletion was restricted to OMP-positive cells. Additionally, C57BL/6J (B6) and OMP-GFP (B6;129P2-*Omp*<sup>tm3(Mom)</sup>/MOM) mice were used.

**Immunohistochemistry.** Perfusion of mice and preparation of mouse olfactory tissues for immunohistochemistry followed previously described methods<sup>34</sup>. Cryosections (10–12 µm) of either human or mouse olfactory tissues were post-fixed using 4% paraformaldehyde in PBS, before blocking and antibody administration. Primary antibodies were: mouse-specific anti- $Na_v1.7$  (1:500, rabbit polyclonal; Millipore), human-specific anti- $Na_v1.7$  (1:500, rabbit polyclonal; Abcam),  $Na_v1.3$  (1:500, rabbit polyclonal; Millipore), OMP (1:3,000, goat polyclonal; gift of F. Margolis), vGluT2 (1:2,000, rabbit polyclonal; Synaptic Systems), tyrosine hydroxylase (TH, 1:3,000, mouse monoclonal; ImmunoStar). Secondary antibodies and conjugated compounds were: Alexa-Fluor 488 donkey-anti-goat (1:1,000; Invitrogen), Alexa-Fluor 555 donkey-anti-rabbit (1:1,000; Invitrogen), Alexa-Fluor 546 Streptavidin (1:200; Invitrogen). Procedures were conducted at room temperature (21 °C), except for incubation in primary antibodies (4 °C). Expression of  $Na_v1.7$  in human was detected by direct immunofluorescence. Expression of  $Na_v1.3$  and  $Na_v1.7$  in mouse was detected by tyramid signal amplification using manufacturer's protocol (TSA-Biotin System, Perkin Elmer). Incubation in primary antibody was for 2–3 days, in biotinylated anti-rabbit antibody (1:400; Jackson ImmunoResearch) for 1 h, in streptavidin-HRP (1:100) for 30 min, in biotinylated tryamid (1:100) for 10 min, and visualized using Alexa 546-conjugated streptavidin (Invitrogen, 1:200). OMP colocalization

was detected using a Alexa 488-conjugated anti-goat secondary antibody. Detection of vGluT2 was exactly as previously described<sup>26</sup>. TH was detected in 30-µm free-floating sections using the avidin-biotin method (Vectastain ABC-Elite, Vector). Incubation in primary TH antibody was for 1 day, in biotinylated horse-anti-mouse secondary antibody (1:400, Vector Laboratories) for 1 h, and in avidin/biotin-HRP complex (Vector) for 90 min. Immunoreactivity was visualized with 0.05 g l<sup>-1</sup> 3,3'-diaminobenzidine and 0.015% H<sub>2</sub>O<sub>2</sub>. Fluorescence images were acquired on either a BX71 microscope attached to a DP71 camera (Olympus) or an LSM 710/ConfoCor-3 microscope (Zeiss). Image stacks are presented as maximum intensity projections, assembled and minimally adjusted in brightness using Adobe Photoshop 6.0.

**Electron microscopy.** Following routine processing for electron microscopy, as previously described<sup>26,35</sup>, thin 70–100-nm sections were cut on a Reichert Ultracut E and examined on a JEOL 1200 transmission electron microscope. Images were captured at ×12,000, digitized at 1,200 dots per inch (DPI), and examined for ultrastructural features of the olfactory sensory axons and their synaptic terminals.

**Electrophysiology.** Whole-cell patch-clamp recordings from individual OSNs were obtained in acute MOE tissue slices of P1–P5 mice<sup>19</sup>. The anterior aspect of the head containing olfactory epithelium and bulb was embedded in agarose (4%), placed in oxygenated, ice-cold extracellular solution (95% O<sub>2</sub>, 5% CO<sub>2</sub>) containing: 120 mM NaCl, 25 mM NaHCO<sub>3</sub>, 5 mM KCl, 5 mM BES (N,N-bis[2-hydroxyethyl]-2-aminoethansulphonic acid), 1 mM MgSO<sub>4</sub>, 1 mM CaCl<sub>2</sub>, 10 mM glucose, osmolality adjusted to 300 mOsm, pH 7.3. Coronal slices (250 µm) were cut on a vibratome (Microm HM 650 V), transferred to a recording chamber and kept under continuous flow (2 ml min<sup>-1</sup>) of oxygenated solution or remained on ice in oxygenated solution until needed (for up to 4 h). Experiments were performed at room temperature. The CsCl-based electrode solution contained: 140 mM CsCl, 1 mM EGTA, 10 mM HEPES, 0.5 mM GTP Na-salt, 2 mM ATP Mg-salt, pH 7.1, 290 mOsm. To assess OSN firing properties under non-invasive conditions, we used extracellular loose-patch recording from OSN knobs<sup>20</sup>. In this case, the septal epithelium of juvenile (P1–P5) or adult mice was dissected and transferred to a recording chamber. Patch pipettes (9–12 MΩ) were filled with a HEPES-based extracellular solution containing: 140 mM NaCl, 5 mM KCl, 1 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 10 mM HEPES, pH 7.4, 300 mOsm. IBMX was prepared in 10 mM stock solution containing 5% dimethylsulphoxide (DMSO) (v/v). For M/T cell recordings, brains were rapidly dissected in ice-cold oxygenated (95% O<sub>2</sub>, 5% CO<sub>2</sub>) solution containing: 83 mM NaCl, 26.2 mM NaHCO<sub>3</sub>, 1 mM NaH<sub>2</sub>PO<sub>4</sub>, 2.5 mM KCl, 3.3 mM MgSO<sub>4</sub>, 0.5 mM CaCl<sub>2</sub>, 70 mM sucrose, pH 7.3, 300 mOsm. Horizontal olfactory bulb slices (300 µm) were cut in this solution. Until use, slices were transferred to oxygenated modified artificial cerebrospinal fluid (ACSF, 95% O<sub>2</sub>, 5% CO<sub>2</sub>) containing: 125 mM NaCl, 25 mM NaHCO<sub>3</sub>, 2.5 mM KCl, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 1 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub> and 25 mM glucose. Recording pipettes had resistances of 4–7 MΩ. M/T cells were identified by size and location of their somata and filled with Lucifer Yellow during patch recording. The intracellular solution contained: 140 mM KCl, 1 mM EGTA, 10 mM HEPES, 1 mM ATP Na-salt, 0.5 mM GTP Mg-salt, 0.1 mM Lucifer Yellow; pH 7.1, 290 mOsm. M/T cells were held at −55 to −60 mV. Input and series resistances were 200–300 MΩ and 15–20 MΩ, respectively. After establishing a whole-cell recording, the ONL was stimulated using a glass electrode (1–1.5 MΩ) filled with HEPES-buffered extracellular solution connected to an electrical stimulator (single stimulus: 20 ms, 40 V, 266–400 µA). The stimulus pipette was placed rostrally to the recorded cell in the ONL. If a given M/T cell showed no postsynaptic response, the position of the stimulus pipette was changed until OSN axon bundles were found that caused M/T cell responses. Ionic currents were analysed using PulseFit 8.54 (HEKA) and IGOR Pro software (Wavemetrics)<sup>36</sup>. OSNs with leak currents >20 pA and M/T cells with leak currents >100 pA (all measured at −70 mV) were excluded from analysis. Cell capacitance (*C<sub>m</sub>*) was monitored using the automated function of the EPC-9 amplifier. A stable *C<sub>m</sub>* value over time was an important criterion for the quality of an experiment. Spike analysis was done off-line using IGOR Pro software with custom-written macros. Chemicals were purchased from Sigma unless otherwise stated. Drugs used in the electrophysiological experiments were prepared as stock solutions in DMSO or distilled water and diluted to the final concentration in HEPES-based extracellular solution. NaCl, MgCl<sub>2</sub>, glucose and CaCl<sub>2</sub> were from Merck. IBMX (100 µM) and cineole (100 µM) were diluted in a HEPES-buffered extracellular solution (<0.1% DMSO) and focally ejected using multibarrel stimulation pipettes.

**Behavioural tests.** The innate olfactory preference test followed previously described procedures<sup>31</sup>. Briefly, mice were habituated to the test conditions before odour exposure. Mice were individually placed in an empty cage for 30 min and then transferred to a new cage. This habituation was repeated three to four times for each animal. Soon after habituation, mice were transferred to the test cage, and a filter paper scented with a test odorant was introduced. Investigation times of the filter paper during the 3-min test period was recorded and quantified. Odour

stimuli were freshly collected male and female B6 mouse urine (5  $\mu$ l), peanut butter (10% w/v, 15  $\mu$ l), milk powder (10% w/v, 15  $\mu$ l), water (15  $\mu$ l) and cineole (100  $\mu$ M, 15  $\mu$ l).

For the innate olfactory avoidance test, following habituation (see innate preference test), a filter paper scented with 5  $\mu$ l TMT (7.6 mM) was placed in one corner of the test cage. Mouse behaviour was recorded for 30 min. The test cage was subdivided into three equally sized areas. Time spent in area 1 of the cage (farthest distance from the TMT source) was evaluated as avoidance, whereas time spent in area 2 (consisting of the TMT source) was evaluated as attraction<sup>37</sup>. Animal movements were tracked with SwisTrack (Swarm Intelligent Systems Group, Swiss Federal Institute of Technology).

For the olfactory habituation–dishabituation assay, following habituation (see innate preference test) mice were exposed for 3 min to distilled water (15  $\mu$ l). This procedure was repeated three times with 1-min intervals, followed by a three-time presentation of female urine (5  $\mu$ l) and a three-time presentation of male urine (5  $\mu$ l). Investigation times during the 3-min test periods were measured.

For the pup retrieval test, lactating mice were habituated to the experiment for several minutes. Experiments were performed in the bedded home cages of the dams. Three pups (1–3-days old) were removed from the nest and randomly

distributed in the cage. The latency for pup retrieval back into the nest was measured. If a dam had not completed retrieval within 10 min the test was terminated, resulting in a latency of 600 s.

Experiments were performed in empty standard cages (38  $\times$  19  $\times$  12 cm) and test substances were applied on filter paper ( $\sim$ 1  $\times$  3 cm). Mouse behaviour was recorded with a digital camera (Sony) for the experimental times indicated. Statistical video analyses were done randomly and blindly. Peanut butter (Barney's Best) and milk powder (Bio-Anfangsmilch, Hipp) were diluted to 10% (w/v) in water.

**IGF-1 assays.** IGF-1 levels were measured by sandwich ELISA (ALPCO Diagnostics). IGF-1 was dissociated from the binding proteins by diluting samples with an acidic buffer. The analytical sensitivity of the assay was 0.029 ng ml<sup>-1</sup>. Inter and intra-assay variability was below 7%. Experiments used plasma of 4–5-weeks-old mice ( $n$  = 4, each genotype).

**Statistics.** Data were analysed using NCSS 2004 statistical software (NCSS). The Student's *t*-test (two-tailed) was used for measuring the significance of difference between two distributions. Multiple groups were compared using a one-way or two-way analysis of variance (ANOVA) with Fisher's LSD as a post hoc comparison. Unless otherwise stated, results are presented as means  $\pm$  s.e.m.